CVCI 2026 Bench2InterActDrive: A Closed-Loop Benchmark for End-to-End Autonomous Driving

Abstract

With the rapid development of end-to-end autonomous driving, there is an increasing need for benchmarks that can evaluate not only general driving capability but also robustness, safety, and generalization under safety-critical and long-tail traffic conditions. Existing benchmarks have provided useful testbeds, but many remain limited by insufficient coverage of realistic corner cases or by evaluation protocols that do not fully capture closed-loop interaction. To address these limitations, we organize the 2026 Benchmark Challenge-ench2InterActDrive for End-to-End Autonomous Driving Evaluation as a challenge session of CVCI 2026.

Built upon the Bench2Drive framework, ench2InterActDrive adopts a closed-loop evaluation protocol and includes 12 categories of challenging scenarios with 144 instances in total, inspired by real-world intelligent driving test cases. It focuses on extreme interactions, multi-agent traffic, and other safety-critical situations that are difficult for current end-to-end systems. To ensure fair and meaningful evaluation, the final results combine the original Bench2Drive score on standard routes with a scenario-aware ench2InterActDrive score on the new extreme interactive scenarios, providing a reproducible and behavior-oriented benchmark for students and researchers.

1 Introduction

End-to-end autonomous driving has become one of the mainstream routes toward large-scale deployment of intelligent driving systems, and is now a major focus of both academic research and industrial development. As end-to-end models continue to advance, there is an increasing need for benchmarks that can systematically evaluate their robustness, safety, and generalization in diverse and safety-critical real-world corner cases. However, their robustness, safety, and generalization under long-tail traffic situations remain far from satisfactory, especially when facing safety-critical interactions, abnormal behaviors of surrounding agents, and complex urban decision-making scenarios. These limitations make comprehensive and fair benchmarking increasingly important for the autonomous driving community.

Existing benchmarks have provided valuable testbeds for evaluating end-to-end driving models, yet they still present several limitations. Open-loop benchmarks are easy to scale, but they cannot fully reflect the consequences of sequential decision-making errors in interactive driving. By contrast, closed-loop simulation can better evaluate complete driving behavior, including recovery ability, safety margins, and interaction handling, which are all crucial for real-world deployment. Therefore, a benchmark specifically designed for closed-loop, safety-critical, and corner-case evaluation is necessary for advancing reliable end-to-end driving research.

To address this need, we organize the 2026 Benchmark Challenge for End-to-End Autonomous Driving Evaluation as a challenge session of CVCI 2026. This benchmark is built upon the Bench2Drive framework and adopts a closed-loop evaluation protocol. It contains 12 categories of representative safety-critical scenarios and 144 scenario instances in total, with scenario design inspired by representative intelligent driving test cases from real-world public evaluations. The benchmark focuses on multi-agent interactions, atypical traffic behaviors, constrained-space maneuvers, and other corner cases that are difficult for current end-to-end systems.

A key design principle of this challenge is that strong performance should not be achieved merely by over-conservative behavior. In particular, a model that simply brakes in most situations may appear safe under some scenario-specific metrics, while still lacking normal route-following capability, traffic efficiency, and general driving competence. For this reason, the official evaluation jointly considers two complementary components: the original Bench2Drive score on standard closed-loop driving routes, and the CVCI scenario score on our newly introduced extreme interactive scenarios. The final leaderboard score is obtained by a weighted combination of these two components, so that submitted models must perform well both in general driving and in safety-critical corner cases.

The main goal of this challenge is not only to compare raw task success rates, but also to systematically evaluate whether a driving model can behave safely, smoothly, and robustly under diverse corner cases. In particular, the challenge emphasizes the following aspects:

1. **Closed-loop evaluation:** models are assessed through full sequential interaction with the simulated environment rather than isolated frame-level prediction.
2. **Scenario diversity:** the benchmark covers multiple safety-critical situations involving vehicles, cyclists, pedestrians, static obstacles, and adverse environmental conditions.
3. **Fairness and reproducibility:** all participants are evaluated under the same simulator configuration, route definitions, scenario triggers, and scoring rules.
4. **Balanced assessment of competence and safety:** the final ranking combines Bench2Drive route-level driving performance with CVCI scenario-aware behavioral evaluation, preventing unrealistic high scores from overly simplistic stop-only strategies.

Through this challenge, we aim to provide a common platform for students and researchers to study the generalization ability of end-to-end driving systems under realistic long-tail conditions, promote transparent comparison among methods, and encourage the development of safer and more robust autonomous driving models.

2 Task and Interface

The task of the 2026 Benchmark Challenge is to develop an end-to-end autonomous driving model that can safely and effectively drive the ego vehicle through a route while handling a variety of safety-critical and corner-case scenarios encountered along the way. Participants are required to submit both the complete end-to-end driving model and the corresponding evaluation materials, including the JSON results file generated by running the official benchmark script and the shell script used for evaluation. The submitted model should be capable of receiving benchmark-defined observations and producing real-time driving outputs for the ego vehicle, such as control commands or planned trajectories.

More specifically, the submitted model should be able to:

- follow the designated route under closed-loop execution,
- maintain safe and lawful driving behavior,
- respond properly to dynamic and unexpected hazards,
- interact reasonably with surrounding traffic participants,

- recover to stable forward driving after the hazardous event is resolved.

The benchmark adopts a unified data interface. During evaluation, the simulator and benchmark framework provide observation streams to the submitted agent at each simulation step. Based on these observations, the agent must output low-level vehicle control commands, typically including steering, throttle, and brake. Participants are not allowed to manually intervene during the official evaluation process. All results are generated automatically under the same runtime protocol.

To ensure fairness, the following principles are enforced:

1. **Fixed benchmark environment:** routes, maps, scenario definitions, weather settings, and evaluation logic are provided by the organizers.
2. **Fixed sensor interface:** all methods receive the same benchmark-defined input modalities.
3. **Fixed runtime setting:** all submissions are evaluated under the same simulation frequency, route execution policy, and scenario activation mechanism.
4. **No manual intervention:** official results are obtained through fully automatic evaluation.

The organizer will provide the necessary benchmark package, including route definitions, scenario files, runtime interface, and evaluation scripts. Participants need to adapt their models to this official interface and ensure that their submission can be executed reproducibly on the organizer's evaluation platform.

3 Platform and Sensors

Built upon the Bench2Drive framework, our benchmark extends the original platform in two important aspects. First, we construct 12 categories of extreme interactive scenarios, resulting in a total of 144 scenario instances, to systematically evaluate end-to-end autonomous driving models under safety-critical and long-tail driving conditions. These scenarios are inspired by challenging real-world intelligent driving tests and emphasize complex dynamic interactions among vehicles, vulnerable road users, and surrounding traffic participants. Second, instead of relying on a single evaluation logic, the benchmark adopts a dual-score protocol: standard Bench2Drive scenarios are evaluated using the original Bench2Drive scoring rule, while the newly introduced CVCI scenarios are evaluated using our scenario-aware scoring rule. The final leaderboard score is then computed by weighted aggregation of the two parts.

The 2026 benchmark challenge is developed on top of the Bench2Drive simulation framework and uses the CARLA simulator as the underlying interactive environment. The benchmark adopts a closed-loop setup, where the tested model continuously receives sensor observations from the simulator and outputs driving actions online. Unlike open-loop playback benchmarks, the ego vehicle's decisions influence subsequent scene evolution, making the evaluation more faithful to real driving behavior and more sensitive to accumulated policy errors.

The simulation platform provides a unified environment for all participants, including predefined maps, routes, weather settings, dynamic traffic agents, and scenario triggers. Each scenario instance is embedded into a route-based evaluation process, so the model is required not only to react to a local hazard, but also to maintain consistent driving performance before, during, and after the hazardous interaction. This design allows the benchmark to evaluate both instantaneous safety response and longer-horizon driving stability.

For sensor input, the benchmark follows the standard Bench2Drive-style embodied driving setup and provides a unified sensor suite for all submitted models. The benchmark is designed primarily for end-to-end driving methods that take raw or minimally processed observations as input. Depending on the challenge track and released interface, the available inputs may include:

- multi-view RGB cameras for front and surrounding perception,
- ego-vehicle state information such as speed and control-related kinematic status,
- route-level navigation information required for high-level driving guidance.

The exact sensor configuration, coordinate definitions, and data formats are fixed by the benchmark toolkit to ensure consistency across all participants. Participants are not allowed to modify the simulator-side sensor deployment during official evaluation. This ensures that performance differences mainly come from algorithm design rather than from privileged sensor settings.

In addition to sensor consistency, the benchmark also fixes key environmental settings such as route files, actor spawning logic, and scenario activation conditions. By controlling these variables, the challenge can provide reproducible and interpretable comparisons among different end-to-end models.

In general, the platform is intended to serve as a practical and scalable testbed for evaluating real driving intelligence in simulation: perception, interaction reasoning, safety-aware planning, constrained maneuvering, and post-conflict recovery are all reflected through closed-loop execution.

4 Scenario Design

The benchmark challenge is conducted within a high-fidelity simulation environment. To evaluate the performance of the coordination controller in extreme conditions, we have designed 12 categories of representative safety-critical scenarios.

High Speed Scenarios

These scenarios emphasize vehicle stability and response during high-velocity maneuvers:

- **Lead Vehicle Occlusion with Abrupt Departure**
- **Lane Blockage under Adjacent Traffic Flow Constraints**
- **Aggressive On-Ramp Merging Interaction**
- **Emergency Avoidance of Lane-Centered Static Obstacle**
- **Narrow Bottleneck Navigation with Disabled Truck**

Urban Scenarios

These scenarios focus on complex multi-agent interactions and vulnerable road user safety:

- **Urban Scenario (General Integrated)**
- **Dense Roundabout Multi-Agent Coordination**
- **Child-Involved Pedestrian Crossing**
- **Urban Disabled Vehicle Interaction**
- **Diagonal Intrusion of Vulnerable Road Users**
- **Sudden Reverse Entry from Lateral Vehicle**
- **Occluded Left-Turn Conflict**

4.1 Lead Vehicle Occlusion with Abrupt Departure

4.1.1 Scenario Description

As illustrated in Figure 1, the scenario takes place on a straight urban road segment and is designed to evaluate the robustness of an autonomous driving system when facing a sudden conflict caused by an abrupt lane-change departure of a lead vehicle that blocks the field of view, in a multi-agent traffic environment. As illustrated in the scenario diagram, the ego vehicle proceeds straight along its lane and maintains car-following status with a lead vehicle that obstructs its forward view. The lead vehicle suddenly changes lanes to the right to avoid a stationary vehicle ahead, leaving the ego vehicle to confront the suddenly exposed hazard. Meanwhile, continuous traffic flow exists on the right side of the ego vehicle, further restricting its ability to perform large-scale evasive maneuvers. The main purpose of this scenario is to assess whether the driving model can properly respond to sudden road hazards. A distinctive feature of this scenario is that the ego vehicle cannot obtain forward road information due to view occlusion, so the hazard exposed after the lead vehicle's sudden lane departure is highly unexpected. This makes the scenario particularly suitable for evaluating risk anticipation, emergency response, as well as decision-making and trajectory planning under spatial constraints.

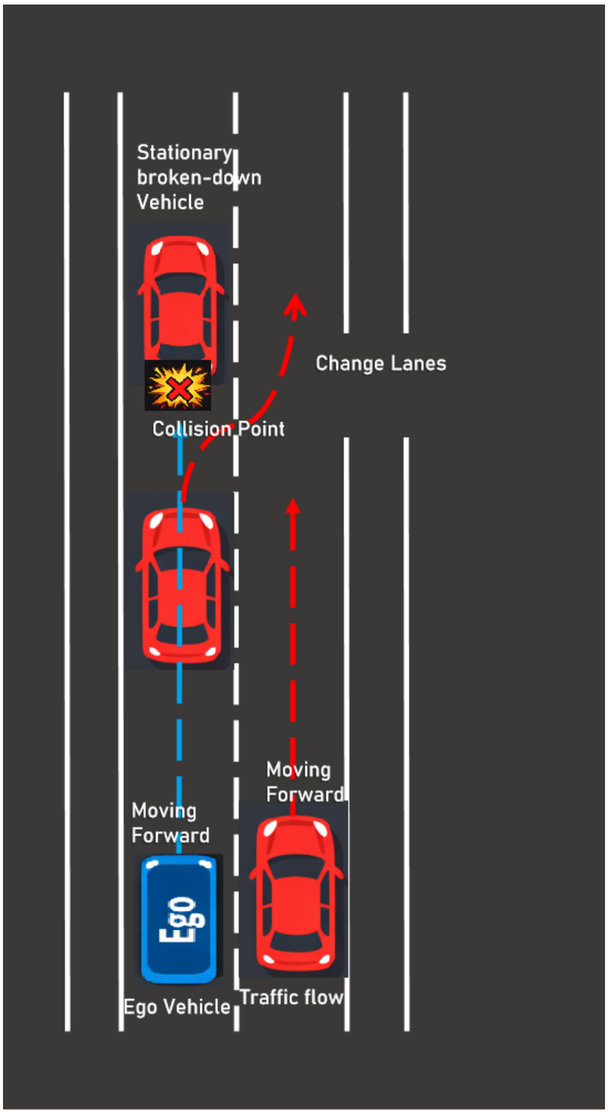

Figure 1: Illustration of the Lead Vehicle Occlusion with Abrupt Departure

4.1.2 Critical Conflict

After the lead vehicle departs, the ego vehicle shall perform graded deceleration and compliant evasive maneuvers in consideration of the constraints from the traffic flow on the right (such as braking smoothly and passing safely when conditions permit). After successfully passing the hazardous area, the ego vehicle is expected to return to a stable car-following or straight-driving state and continue along its planned route. In terms of evaluation, this scenario assesses not only the immediate safety response of the ego vehicle to suddenly exposed visual risks, but also its accuracy in risk anticipation under view occlusion and information delay, as well as its ability to quickly resume normal driving behavior from a sudden conflict in a multi-vehicle constrained environment.

4.1.3 Correct Handling Procedure

After responding to the reversing vehicle and safely passing the conflict area, the ego vehicle is expected to return to a stable forward-driving state and continue along its intended route. In this sense, the scenario evaluates not only the immediate safety response to a sudden hazard, but also the vehicle's ability to recover normal driving behavior after the conflict has been resolved.

4.2 Lane Blockage under Adjacent Traffic Flow Constraints

4.2.1 Scenario Description

As illustrated in Figure 2, this scenario is set on a highway segment and is designed to evaluate the robustness of an autonomous driving system when facing temporary construction closure and sudden lane narrowing on the right lane, under the constraint of dense traffic flow in the adjacent left lane, in a multi-agent traffic environment. As illustrated in the scenario diagram, the ego vehicle is driving normally along the right lane of the highway, with dense and fast-moving traffic on the left, and no obvious gaps for lane changing. A temporary construction zone is set ahead on the right lane with traffic barrels (cones), causing sudden lane narrowing and forming a driving blockage. During driving, the ego vehicle continuously attempts to change lanes to the left to avoid the construction area, but due to the constraint of dense traffic on the left, multiple lane change attempts fail. It is not until the vehicle approaches the cones and guide signs that it initiates emergency braking to reduce speed, and finally collides with the cones (preset conflict result of the scenario). The main purpose of this scenario is to assess the decision-making, planning, and emergency response capabilities of the driving model under the dual pressure of "sudden road blockage + dynamic constraints of adjacent lanes". A distinctive feature of this scenario is that the hazard (road blockage) has clear guide prompts, but the dynamic traffic flow in the adjacent lane forms a rigid constraint, making it impossible for the ego vehicle to avoid the risk through conventional lane changing. The ego vehicle needs to quickly switch to emergency disposal mode after "failed lane changing". Therefore, this scenario is particularly suitable for evaluating the system's dynamic risk anticipation, rationality of lane change decisions, and the timing and intensity control of emergency braking.

4.2.2 Critical Conflict

The critical conflict occurs at the moment when the ego vehicle fails to change lanes to the left multiple times, approaches the cones and guide signs, and is about to collide with the construction closure area. This interaction is driven by three coupled factors: first, the sudden road blockage caused by temporary construction on the right lane, which directly blocks the original driving path of the ego vehicle and forms a rigid collision risk; second, the dense and fast-moving traffic in the left lane, which constitutes a dynamic spatial constraint, making it impossible for the ego vehicle to avoid the risk by changing lanes to the left and losing the core evasion path; third, the insufficient lane change decision-making and timing control of the ego vehicle, which fails to complete the lane change within an effective distance, and the emergency braking is initiated too late to completely avoid the collision. Unlike sudden conflict scenarios on urban roads, the core conflict of this scenario lies in "predictable hazard + unavoidable dynamic constraints". It is not merely an emergency braking scenario, but requires the ego vehicle to quickly complete the entire decision-making process of "construction risk identification → lane change feasibility assessment → lane change attempt → lane change failure → emergency braking". It is necessary to take into account the dynamic changes of traffic flow on the left, as well as accurately control the braking timing, to avoid collisions caused by late braking or rear-end collisions caused by sudden braking.

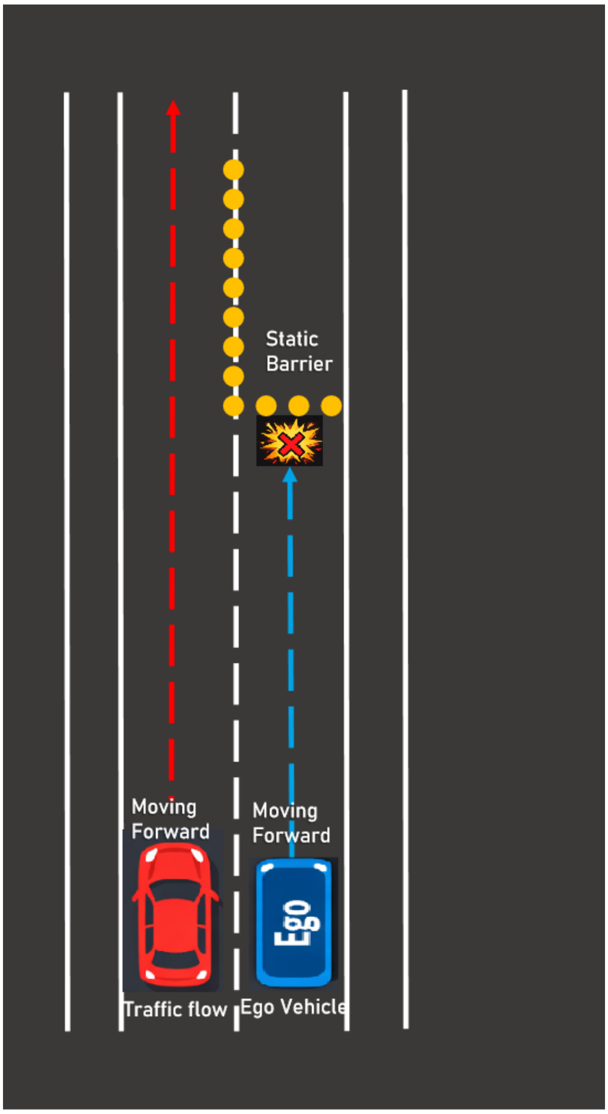

Figure 2: Illustration of the Lane Blockage under Adjacent Traffic Flow Constraints.

4.2.3 Correct Handling Procedure

After identifying the construction guide signs and cones ahead on the right lane and confirming the lane narrowing and blockage, the ego vehicle should immediately initiate risk assessment, and attempt to change lanes to the left in combination with the traffic density and driving gaps on the left (prioritize gaps with sufficient safe distance to avoid forced lane changing); if multiple lane change attempts fail due to dense traffic on the left, it is necessary to initiate graded deceleration in advance (instead of emergency braking only when approaching the cones), gradually reduce the speed, and continuously observe the traffic gaps on the left to seize the opportunity to complete lane change and evasion. If lane change still cannot be completed before reaching the cones, the ego vehicle should perform smooth emergency braking to reduce the speed to a safe range, and minimize the collision force with the cones (reduce vehicle and facility damage); after the collision (if it occurs), the ego vehicle should immediately turn on the hazard warning flashers, and after confirming the safety of surrounding traffic flow, slowly drive away from the conflict area and resume normal driving (if normal driving is not possible, it should park in the emergency lane in a standardized manner and send a distress signal). In terms of evaluation, this scenario assesses not only the ego vehicle's ability to identify and anticipate highway construction blockages, its ability to dynamically perceive left traffic flow and make lane change decisions, but also focuses on verifying its emergency braking control after failed lane changes, as well as its risk disposal and driving recovery capabilities after collisions.

4.3 Aggressive On-Ramp Merging Interaction

4.3.1 Scenario Description

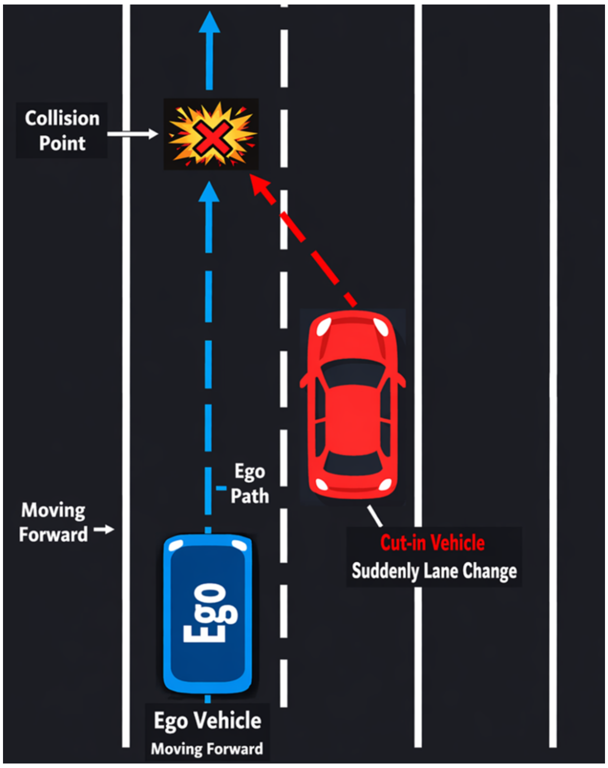

Figure 3: Illustration of the Aggressive On-Ramp Merging Interaction.

As illustrated in Figure 3, the scenario takes place on a high-speed highway segment and is designed to evaluate the robustness of an autonomous driving system when encountering an aggressive cut-in maneuver from a neighboring vehicle under dense traffic conditions. In this scenario, the ego vehicle travels forward at highway speed along its current lane, maintaining normal longitudinal motion. Meanwhile, a vehicle in the adjacent lane suddenly steers into the ego vehicle's lane with insufficient lateral clearance and limited temporal gap, creating an immediate collision risk in front of the ego vehicle. At the same time, surrounding traffic in adjacent lanes restricts the ego vehicle's ability to perform a large-scale evasive maneuver, forcing the system to make a rapid and safe decision under tight dynamic constraints. The main purpose of this scenario is to assess whether the driving model can properly handle highly aggressive and socially non-compliant driving behavior that frequently occurs in real-world highway traffic. Unlike ordinary lane-changing situations, the intruding vehicle does not wait for a sufficient safety gap before merging, but instead forces its way into the ego vehicle's lane in a short time window. This makes the scenario particularly suitable for evaluating risk anticipation, short-horizon motion prediction, emergency response, and decision-making in high-speed interactive traffic environments.

4.3.2 Critical Conflict

The critical conflict arises at the moment when the intruding vehicle crosses the lane boundary and enters the ego vehicle's intended forward corridor. This interaction is governed by the coupling of three key factors: the ego vehicle's high-speed longitudinal motion, the rapid lateral intrusion of the cut-in vehicle, and the limited maneuvering space caused by surrounding traffic conditions. A reliable driving system must determine whether to brake, slightly adjust its lateral positioning, or maintain stability while ensuring safety and ride comfort.

4.3.3 Correct Handling Procedure

Once the intruding vehicle has completed the merge and the immediate danger has been mitigated, the ego vehicle should gradually resume stable forward driving and continue along its intended route. In this sense, the scenario evaluates not only the immediate response to a sudden highway intrusion, but also the system's ability to recover normal driving behavior after a high-speed interactive conflict.

4.4 Emergency Avoidance of Lane-Centered Static Obstacle

4.4.1 Scenario Description

As illustrated in Figure 4, the scenario takes place on a straight two-lane road segment and is designed to evaluate an autonomous vehicle's response to a static, multi-lane obstruction. As illustrated in the scenario diagram, the ego vehicle proceeds straight along the left lane. Ahead, a disabled vehicle is statically positioned at a 45-degree angle across the central lane divider, obstructing both the ego vehicle's current lane and the adjacent right lane, creating an immediate and severe path occlusion

4.4.2 Critical Conflict

The critical conflict arises directly in the ego vehicle's forward path due to the angled blockage. This interaction is shaped by the overlap of two primary factors: the ego vehicle's straight longitudinal motion and the complex spatial constraint imposed by the disabled vehicle simultaneously occupying parts of both the left and right lanes. While a simple lane change is not viable, a narrow drivable corridor may exist, forcing the system to make a critical choice: attempt to pass to the left or to the right of the disabled vehicle. The scenario is thus not just about avoidance, but is a strict test of the system's high-fidelity perception, precise distance estimation, and complex path-finding algorithms to identify and select the safest narrow path or execute an immediate emergency braking maneuver if no path is viable.

4.4.3 Correct Handling Procedure

Upon detecting the angled disabled vehicle, the ego vehicle is expected to execute a safe, controlled, and timely emergency stop before reaching the collision point, recognizing that lateral evasion is obstructed. After successfully coming to a complete halt at a safe distance, the ego vehicle should continuously assess the surrounding environment and remaining drivable space. If a safe micro-routing or narrow traversal maneuver eventually becomes viable, it may carefully navigate around the hazard; otherwise, it must maintain a safe stationary state until the blockage is cleared. In this sense, the scenario evaluates not only the immediate critical braking response to an unavoidable hazard, but also the system's post-stop strategic decision-making in a heavily restricted traffic situation.

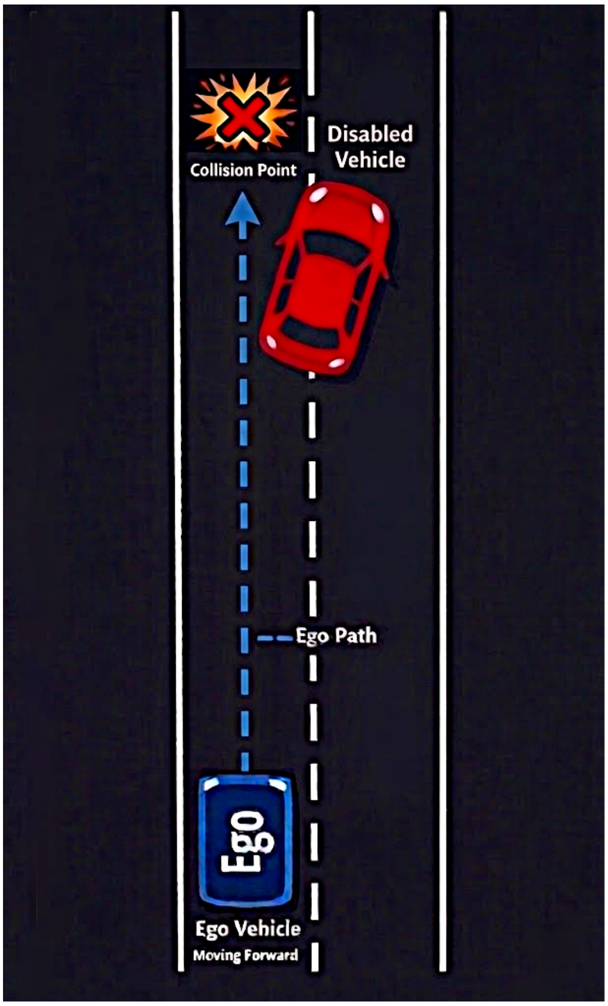

Figure 4: Illustration of the Emergency Avoidance of Lane-Centered Static Obstacle.

4.5 Narrow Bottleneck Navigation with Disabled Truck

4.5.1 Scenario Description

As illustrated in Figure 5, during highway driving, a vehicle approaches a construction zone marked by a row of traffic cones. Simultaneously, a large truck is parked stationary along the roadside. Between the row of traffic cones and the parked heavy truck, there is a narrow passage measuring exactly 2.55 meters in width. This gap is physically wide enough to accommodate a standard passenger vehicle, provided it passes through at a low speed.

4.5.2 Critical Conflict

While the 2.55-meter passage is technically passable, it leaves a severely restricted margin for error, particularly under poor nighttime visibility. The core conflict is that the vehicle (or its smart driving/AEB system) must accurately identify the narrow gap, completely disengage high-speed cruising, and significantly decelerate to a crawling speed before entering the channel. If the driver or the automated system fails to decelerate in time, misjudges the width, or attempts to pass at highway speeds, a severe collision with the heavy truck or the construction barricades is inevitable.

4.5.3 Correct Handling Procedure

To safely navigate this scenario, the vehicle or driver must first utilize high beams or advanced LiDAR/vision sensors to identify the reflective traffic cones, the stationary heavy truck, and the exact 2.55-meter passable gap as early as possible. Upon detection, it is imperative to immediately disengage high-speed cruise control and apply the brakes proactively, reducing the vehicle from highway speeds to a very slow, manageable crawl well before reaching the choke point. Once decelerated, the vehicle must be meticulously aligned to the absolute center of the narrow passage to ensure equal clearance between the cones on the left and the truck on the right. Finally, the system must maintain a steady, cautious speed while navigating through the channel, continuously monitoring the side mirrors to prevent side-scrapes and remaining highly alert for secondary hazards or road workers until completely clear of the construction zone.

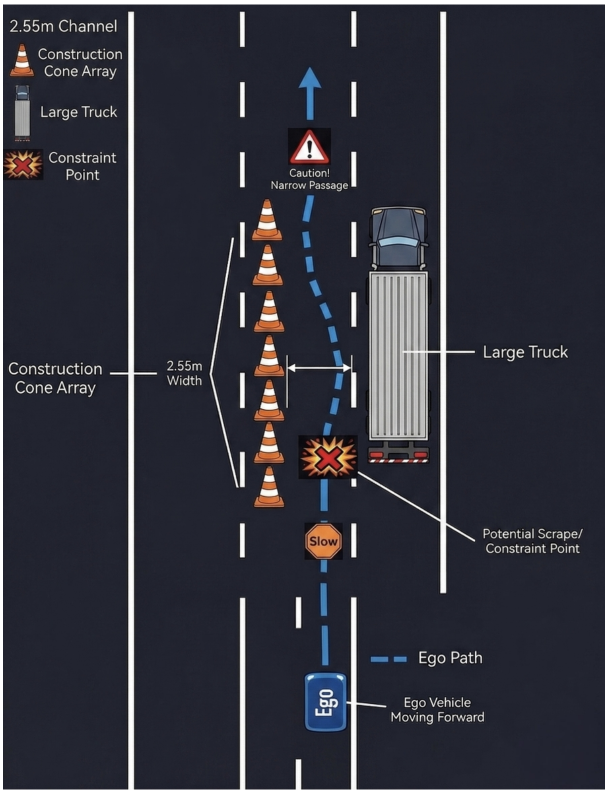

Figure 5: Illustration of the Narrow Bottleneck Navigation with Disabled Truck.

4.6 Dense Roundabout Multi-Agent Coordination

4.6.1 Scenario Description

As illustrated in Figure 6, this scenario is set in an urban road roundabout merging zone, designed to evaluate the decision-making robustness of autonomous driving systems under the intertwining of complex geometric structures and multi-directional dynamic traffic flows. As shown in the scenario schematic, the ego vehicle is attempting to merge into a large-scale roundabout from a straight approach lane. The challenge lies in the superposition of compound obstacles: first, a stationary obstacle vehicle occupies part of the lane at the merge entrance; second, a high-speed oncoming vehicle is passing through the outer ring of the roundabout; finally, there is a continuous fleet with right-of-way on the inner ring of the roundabout. The core purpose of this scenario is to assess the comprehensive processing capability of the autonomous driving model in situations with limited perception and overlapping right-of-way priorities. Unlike ordinary single merging scenarios, the ego vehicle must bypass the stationary obstacle while accurately capturing the dynamics of the high-speed oncoming vehicle from the side and rear, and strictly abide by roundabout traffic rules (inner ring priority). This is critical for evaluating the system's lateral perception flexibility in narrow spacing, longitudinal speed control, and understanding of complex right-of-way rules.

4.6.2 Critical Conflict

The key conflicts in this scenario consist of interactions across three dimensions: the obstacle vehicle at the merge entrance not only compresses the maneuvering space of the ego vehicle but may also create a blind spot for side and rear perception due to deviations in the vehicle body angle; when the ego vehicle bypasses the obstacle to merge into the roundabout, it will form a spatio-temporal conflict with the accelerating oncoming vehicle from the outer ring on the side and rear, which may lead to collisions in case of insufficient prediction or limited avoidance space; after entering the roundabout, the ego vehicle must comply with right-of-way rules, wait in a restricted space for the inner ring fleet to pass, and shall not cut in forcibly.

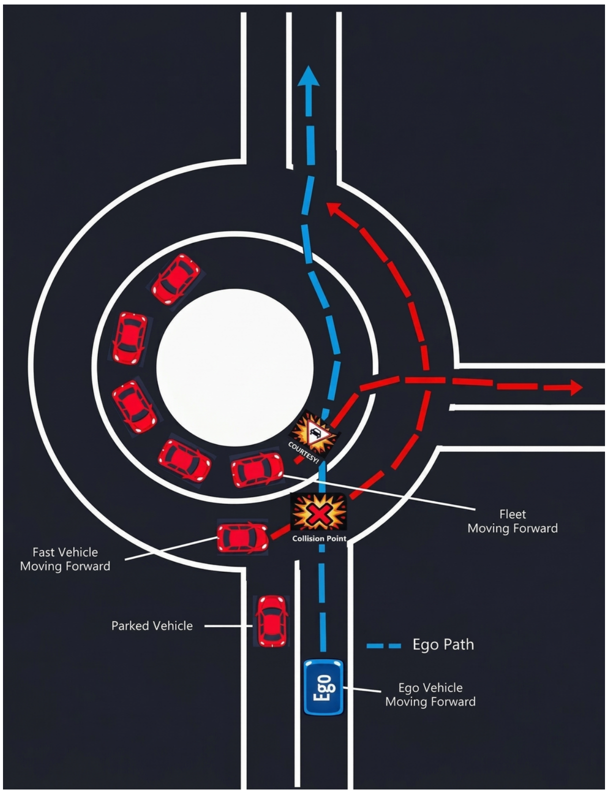

Figure 6: Illustration of the Dense Roundabout Multi-Agent Coordination.

4.6.3 Correct Handling Procedure

As the ego vehicle approaches the obstacle at the merge entrance, it shall slow down in advance, adjust the cornering angle, keep a proper distance from the obstacle, and reserve sufficient avoidance margin; before and during merging, continuously monitor oncoming vehicles from the side and rear, and if an outer ring vehicle is found accelerating, slow down or stop to yield without forced cutting; after entering the roundabout, identify the inner ring fleet, stop smoothly and yield unconditionally until the entire inner ring fleet has passed; after confirming no safety threats, accelerate smoothly to merge into the target lane and resume normal cruising.

4.7 Child-Involved Pedestrian Crossing

4.7.1 Scenario Description

As illustrated in Figure 7, this scenario is set on a straight urban road with two same-direction lanes and is designed to evaluate the defensive driving capabilities of an autonomous driving system when facing a dynamic blind spot. The ego vehicle proceeds straight in its lane, while two forward-moving e-bikes on the right side create a moving visual obstruction. As the ego vehicle approaches, four elementary students suddenly emerge from in front of the moving e-bikes, crossing the ego vehicle's path transversally in a structured, parallel formation. The primary purpose of this scenario is to evaluate the defensive driving capabilities of an autonomous driving system in complex, occluded urban environments. It specifically tests the perception system's ability to detect and classify multiple small, unpredictably moving targets (children) at an extremely close range, as well as the decision-making system's emergency response strategy when facing sudden lateral intrusions from a visual blind spot.

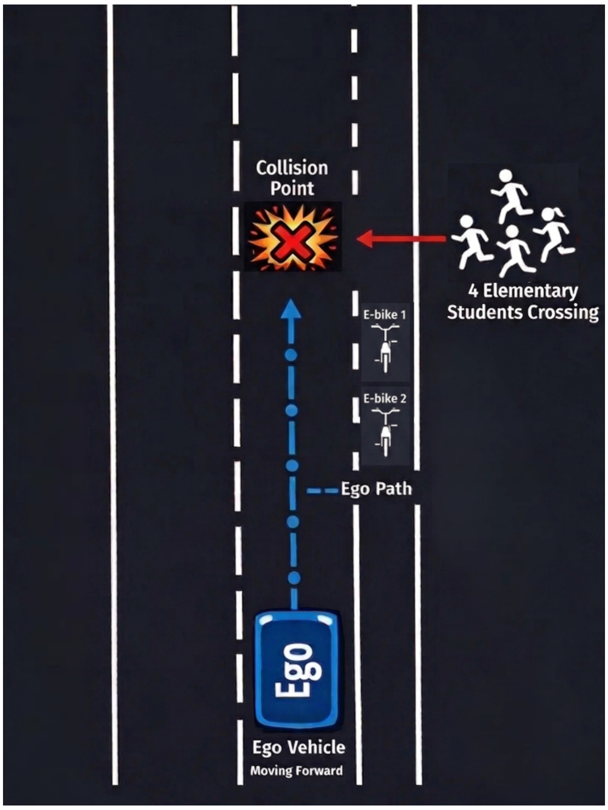

Figure 7: Illustration of the Child-Involved Pedestrian Crossing.

4.7.2 Critical Conflict

The critical conflict arises the exact moment the structured formation of children emerges from visual cover and enters the ego vehicle's path. This interaction is shaped by the overlap of the ego vehicle's straight longitudinal speed and the sudden, uniform lateral intrusion of the children, making the scenario a strict test of perception, motion prediction, risk assessment, and decision-making under constrained time. The system must reliably detect and predict the motion of the entire structured group, as their motion is uniform and non-random.

4.7.3 Correct Handling Procedure

Upon detecting the structured formation of children emerging from the dynamic blind spot, the ego vehicle is expected to immediately classify the group, quickly assess the severe collision risk, and execute a full-force emergency braking maneuver to bring the vehicle to a complete and stable halt before the collision point. Since an evasive lane change is not a predictable or safe option given the suddenness of the hazard, maximum braking is the essential safety response. After successfully avoiding the collision and remaining stationary to ensure all children have safely cleared the drivable area, the ego vehicle is expected to release the brakes, return to a stable forward-driving state, and safely continue along its intended route, demonstrating both emergency hazard mitigation and post-conflict recovery.

4.8 Urban Disabled Vehicle Interaction

4.8.1 Scenario Description

As illustrated in Figure 8, the scenario takes place on a multi-lane urban road segment near a diverging exit and is designed to evaluate the decision-making and trajectory planning capabilities of an autonomous driving system when its path is obstructed under tight spatial constraints. As illustrated, the ego vehicle is proceeding straight in the second lane from the left when a stationary disabled vehicle, preceded by a warning sign, blocks its travel corridor and creates an immediate collision risk. Because three vehicles are traveling in parallel on the right side, a rightward evasive maneuver is impossible, forcing the ego vehicle to perform a complex sequence: temporarily merging into the leftmost lane—which is designated as an exit ramp—to bypass the hazard before immediately returning to its original straight-ahead lane. The primary purpose of this scenario is to assess the driving model's ability to execute multi-stage maneuvers under severe spatial constraints and conflicting lane semantics. Unlike standard obstacle avoidance, this situation forces the system to utilize a lane intended for "exiting" while resisting the command to actually depart the main road, testing the system's ability to prioritize short-term emergency bypasses without compromising long-term navigational goals. Consequently, the scenario is particularly effective for evaluating high-precision trajectory planning and the logic required to manage dynamic constraints from adjacent traffic while navigating around static obstacles in a complex road environment.

4.8.2 Critical Conflict

The critical conflict arises the exact moment the structured formation of children emerges from the visual cover of the moving e-bikes and enters the ego vehicle's forward path. This interaction is shaped by the overlap of three coupled factors: the ego vehicle's straight longitudinal speed, the forward motion of the occluding e-bikes, and the sudden, uniform lateral intrusion of the children. The system must reliably detect and predict the motion of the entire structured group in extremely limited time, as their sudden crossing leaves virtually no room for safe lateral evasion.

4.8.3 Correct Handling Procedure

The critical conflict arises the exact moment the structured formation of children emerges from the visual cover of the moving e-bikes and enters the ego vehicle's forward path. This interaction is shaped by the overlap of three coupled factors: the ego vehicle's straight longitudinal speed, the forward motion of the occluding e-bikes, and the sudden, uniform lateral intrusion of the children. The system must reliably detect and predict the motion of the entire structured group in extremely limited time, as their sudden crossing leaves virtually no room for safe lateral evasion.

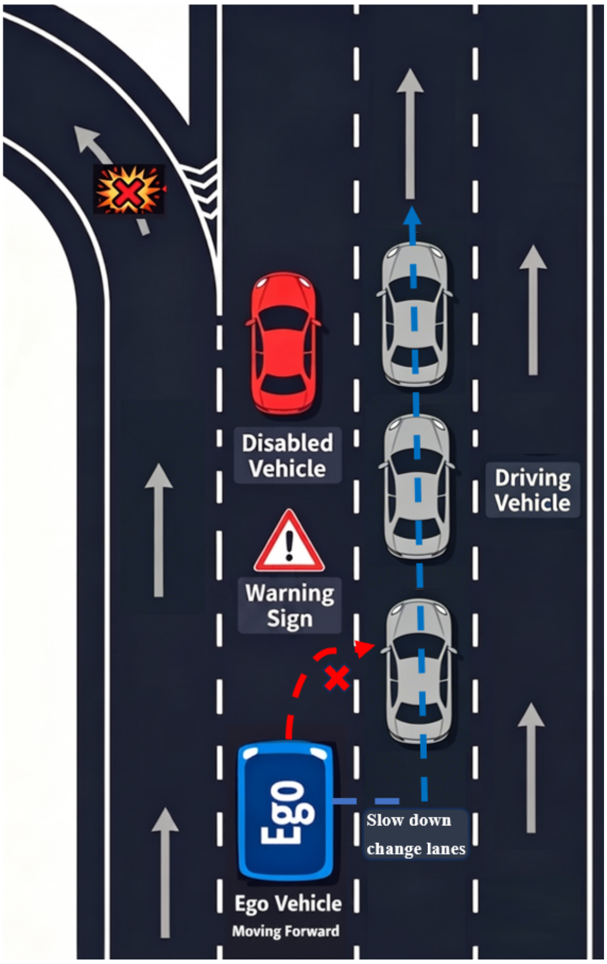

Figure 8: Illustration of the Urban Disabled Vehicle Interaction.

4.9 Diagonal Intrusion of Vulnerable Road Users

4.9.1 Scenario Description

As illustrated in Figure 9, this scenario takes place at an urban intersection and is designed to evaluate the robustness of an autonomous driving system when facing consecutive conflicts during a left turn: a sudden e-bike crossing followed by a pedestrian crossing. As illustrated in the scenario diagram, the ego vehicle enters the intersection at a certain speed to make a left turn. At this moment, an e-bike suddenly appears from the left side and travels diagonally across the intersection, directly cutting across the ego vehicle's left-turn path, creating the first critical conflict. As the ego vehicle approaches the completion of the left turn and is about to enter the left-turn lane, a pedestrian appears on the crosswalk within the ego vehicle's travel path. Due to potential visual occlusion caused by the previous e-bike event, the pedestrian crossing constitutes a second conflict. The main purpose of this scenario is to evaluate the robustness of an autonomous driving system in consecutive, dynamic urban traffic conflicts. Unlike a single emergency event, this scenario involves two temporally close and qualitatively different risk events (lateral e-bike crossing + longitudinal pedestrian crossing). It is particularly suitable for assessing the system's capabilities in multi-object tracking, risk prioritization, decision-making coherence, and emergency braking with smooth recovery in complex intersection environments.

4.9.2 Critical Conflict

The critical conflict emerges in two consecutive phases. In the first phase, the e-bike cuts diagonally into the left-turn path. Due to its atypical trajectory (lateral crossing rather than same-direction or oncoming traffic), the ego vehicle, having just entered the intersection, struggles to anticipate the hazard in time, facing a risk of emergency braking or collision. In the second phase, after the ego vehicle has avoided the e-bike and is about to complete the left turn, a pedestrian appears in the ego vehicle's travel path. Due to potential visual occlusion caused by the e-bike event, the ego vehicle faces a secondary collision risk. Therefore, this scenario is not a single-obstacle response problem but a coupled challenge involving sequential decision-making, recovery from partial occlusion, risk transition recognition, and path continuity maintenance.

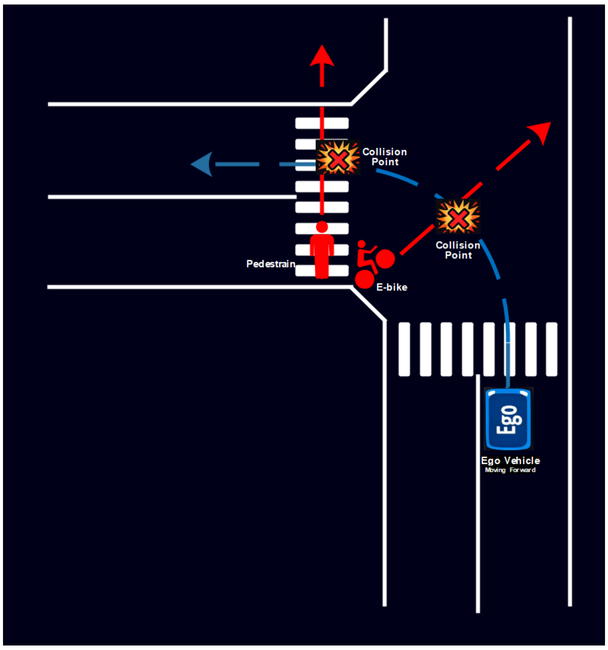

Figure 9: Illustration of the Diagonal Intrusion of Vulnerable Road Users.

4.9.3 Correct Handling Procedure

After entering the intersection, the ego vehicle decelerates upon detecting the e-bike in its field of view, then brakes when a certain distance from the pedestrian and waits for the pedestrian to safely clear its travel path.

4.10 Sudden Reverse Entry from Lateral Vehicle

4.10.1 Scenario Description

As illustrated in Figure 10, the scenario takes place on a straight urban road segment and is designed to evaluate the robustness of an autonomous driving system when facing a sudden reverse-entry conflict caused by a laterally parked vehicle in a multi-agent traffic environment. As illustrated in the scenario diagram, the ego vehicle proceeds straight along its lane, while a parked vehicle on the right side unexpectedly reverses sideways into the ego's travel corridor, creating an immediate collision risk. Meanwhile, a cyclist and another forward-moving vehicle occupy the adjacent space on the left, reducing the ego vehicle's freedom to perform a large evasive maneuver.

The main purpose of this scenario is to assess whether the driving model can properly respond to abnormal roadside vehicle behavior that deviates from normal traffic expectations.

Unlike standard cut-in or crossing scenarios, the hazardous vehicle is initially static and positioned outside the ego lane. This makes the scenario particularly suitable for evaluating risk anticipation, emergency response, and decision-making under constrained space.

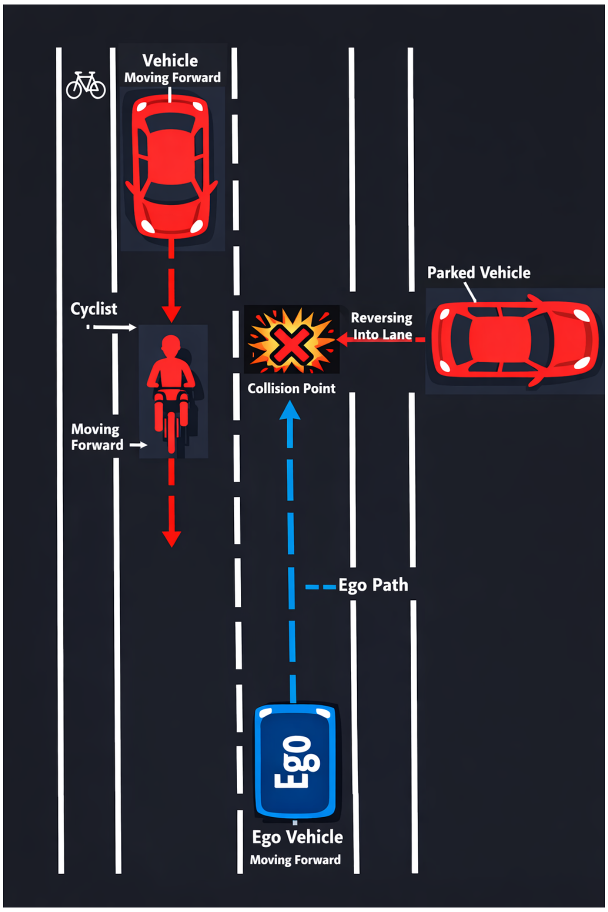

Figure 10: Illustration of the Sudden Reverse Entry from Lateral Vehicle scenario.

4.10.2 Critical Conflict

The critical conflict arises at the point where the reversing vehicle enters the ego vehicle's forward path. This interaction is shaped by the overlap of three coupled factors: the ego vehicle's straight longitudinal motion, the reverse lateral intrusion of the parked vehicle, and the spatial constraints imposed by the cyclist and the adjacent moving vehicle. Therefore, the scenario is not merely an emergency braking case, but a coupled interaction problem that requires perception, motion prediction, risk assessment, and constrained trajectory planning.

4.10.3 Correct Handling Procedure

After responding to the reversing vehicle and safely passing the conflict area, the ego vehicle is expected to return to a stable forward-driving state and continue along its intended route. In this sense, the scenario evaluates not only the immediate safety response to a sudden hazard, but also the vehicle's ability to recover normal driving behavior after the conflict has been resolved.

4.11 Irregular Trajectory of Electric Two-Wheelers

4.11.1 Scenario Description

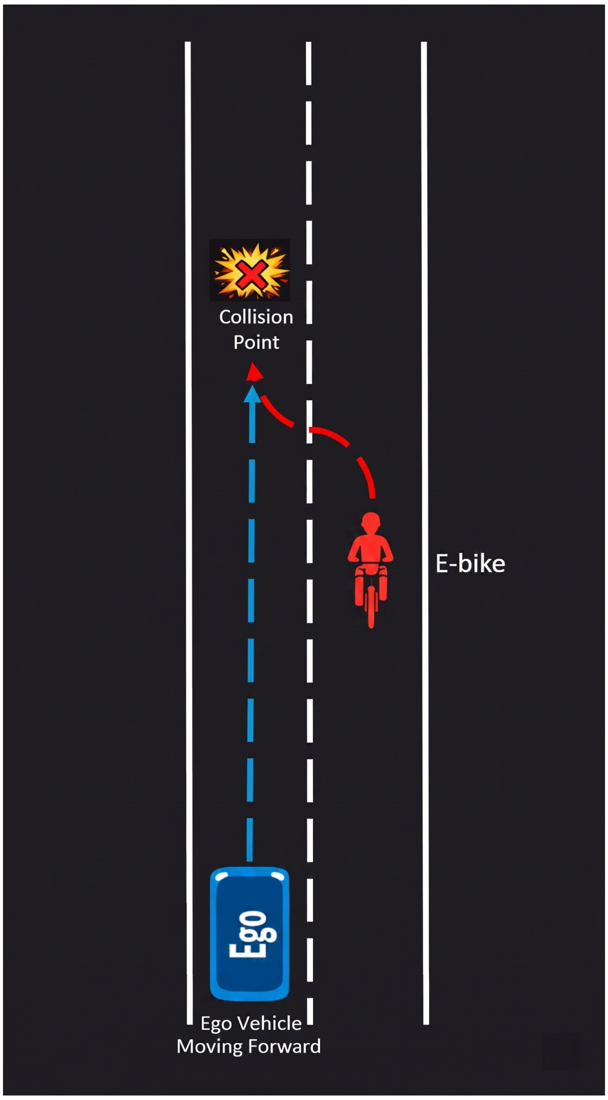

Figure 11: Illustration of the Irregular Trajectory of Electric Two-Wheelers.

As illustrated in Figure 11, this scenario takes place on a straight, two-lane urban road segment and is designed to evaluate the robustness of an autonomous driving system when facing a sudden cut-in conflict caused by a highly maneuverable vulnerable road user (VRU). The ego vehicle proceeds forward at a steady speed along the left lane, while an e-bike traveling in the adjacent right lane unexpectedly and sharply steers laterally into the ego vehicle's travel corridor. This sudden intrusion creates an immediate frontal collision risk. The main purpose of this scenario is to assess whether the driving model can properly anticipate and respond to dynamic, aggressive maneuvers by micromobility vehicles that abruptly deviate from stable lane-keeping behavior.

4.11.2 Critical Conflict

The critical conflict arises at the exact point where the cutting-in e-bike enters the ego vehicle's forward path. This interaction is shaped by the overlap of two primary factors: the ego vehicle's steady longitudinal motion and the sudden lateral intrusion of the e-bike. Because e-bikes can change direction much more rapidly than standard vehicles and often travel at different speed profiles, this scenario is a strict test of the system's high-fidelity perception, instantaneous trajectory prediction, and real-time risk assessment to prevent a rear-end or side-swipe collision.

4.11.3 Correct Handling Procedure

Upon detecting the aggressive cut-in maneuver, the ego vehicle is expected to immediately evaluate the severe collision risk and execute a decisive and smooth emergency deceleration. If the e-bike fully enters and establishes itself in the lane, the ego vehicle must dynamically adjust its speed to establish and maintain a safe following distance. After successfully mitigating the immediate hazard and stabilizing the situation, the ego vehicle is expected to return to a normal forward-driving state and safely continue along its intended route. This evaluates not only the critical braking and yielding response but also the system's ability to seamlessly recover standard driving behavior post-conflict.

4.12 Occluded Left-Turn Conflict

4.12.1 Scenario Description

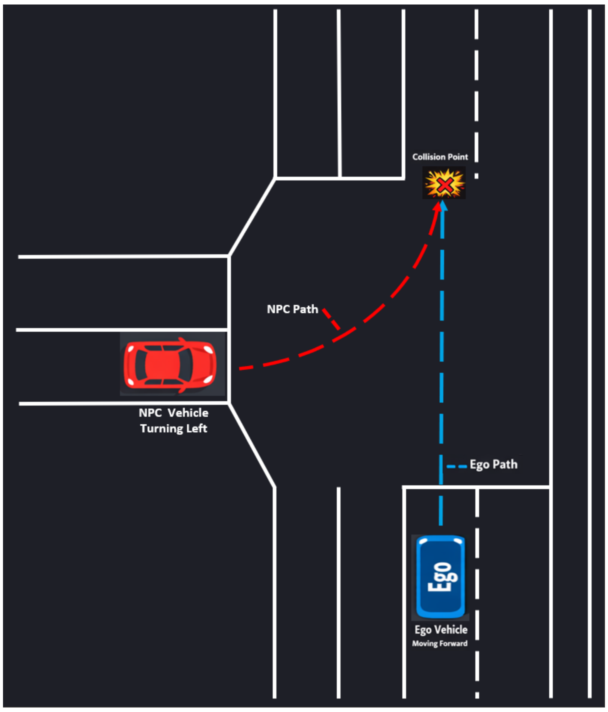

Figure 12: Illustration of the Occluded Left-Turn Conflict.

As illustrated in Figure 12, the scenario takes place at a T-intersection in an urban environment and is designed to evaluate the responsiveness of an autonomous driving system under limited visibility conditions when facing a high-speed cut-in conflict caused by a vehicle from a lateral roadway. As illustrated in the scenario diagram, the ego vehicle proceeds straight along the main road and enters the intersection. Meanwhile, another vehicle from the perpendicular road performs a left-turn maneuver and aggressively merges into the ego vehicle's target lane without yielding. Due to occlusions caused by roadside structures, the approaching vehicle from the left remains partially hidden before entering the intersection, making its behavior highly sudden and unpredictable, thereby creating a significant collision risk. The primary objective of this scenario is to assess whether the driving model can promptly detect unexpected high-risk dynamic agents and respond appropriately. This scenario represents a classic high-difficulty cut-in case, making it suitable for evaluating the system's risk perception, emergency response, and decision-making capabilities under constrained visibility conditions.

4.12.2 Critical Conflict

The critical conflict occurs at the intersection point where the left-side vehicle intrudes into the ego vehicle's forward path. This interaction is shaped by the coupling of several factors: the ego vehicle's continuous longitudinal motion, the aggressive and non-compliant intrusion of the left-turning vehicle, and the visibility constraints introduced by road structures or barriers. Therefore, this scenario is not merely an emergency braking case, but a coupled interaction problem involving uncertain perception, motion prediction, risk assessment, and constrained trajectory planning.

4.12.3 Correct Handling Procedure

Upon detecting the fast-approaching vehicle intruding from the left, the ego vehicle is expected to decelerate in a timely manner and yield to avoid potential collision. After the conflict is resolved and the intersection is safely cleared, the ego vehicle should return to a stable forward-driving state and continue along its intended route. In this sense, the scenario evaluates not only the system's immediate safety response to sudden hazards, but also its ability to recover normal driving behavior after the conflict has been resolved.

5 Evaluation Metrics

A central goal of the 2026 benchmark challenge is to provide a more informative evaluation than simple success-or-failure statistics. Since end-to-end autonomous driving is a sequential decision-making problem, a model should be judged not only by whether it eventually reaches the destination, but also by how it behaves throughout the driving process. Crucially, as CVCI Bench2InterActDrive focuses on safety-critical and extreme interactive scenarios, a significant challenge arises: many of these scenarios might inadvertently assign high safety scores to an agent that simply remains stationary. To prevent such degenerate behaviors—where models avoid complex interactions through overly conservative braking or stop-only policies—the benchmark adopts a dual-score evaluation protocol: standard Bench2Drive routes are scored with the original Bench2Drive metric, while Bench2InterActDrive extreme interactive scenarios are scored with our scenario-aware metric. The final leaderboard score is the weighted sum of these two parts.

5.1 Bench2Drive Score on Standard Routes

For the original Bench2Drive scenarios, we keep the native route-level evaluation rule. Let R_j denote the route completion score of the j-th Bench2Drive route, and let P_j^{inf} denote the cumulative infraction penalty on that route. The Bench2Drive route score is defined as:

$$S_j^{\mathrm{B2D}} = R_j \times P_j^{\mathrm{inf}}$$

where:

- $R_j \in [0, 100]$ measures how much of the route is successfully completed;
- $P_j^{\mathrm{inf}} \in [0, 1]$ is the product of penalty factors induced by traffic infractions such as collisions, route deviations, red-light violations, stop-sign violations, and other unsafe behaviors.

The overall Bench2Drive score is then computed as the average over all N_{B2D} standard routes:

$$\overline{S}^{\text{B2D}} = \frac{1}{N_{\text{B2D}}} \sum_{j=1}^{N_{\text{B2D}}} S_j^{\text{B2D}}$$

This term ensures that submitted models retain normal closed-loop driving competence, including route following, traffic-rule compliance, and stable driving over standard benchmark routes.

5.2 Bench2InterActDrive Score on Extreme Interactive Scenarios

For the newly introduced Bench2InterActDrive extreme interactive scenarios, we use a structured scenario-aware metric that explicitly evaluates whether the agent performs the intended key behaviors in each safety-critical situation. For the i-th Bench2InterActDrive scenario, the score is defined as:

$$S_i^{\text{B2IAD}} = \text{BaseScore}_i \times \text{Penalty}_i$$

Here, BaseScore_i measures the completion quality of the scenario-specific intended behaviors, while Penalty_i is a collision-based penalty coefficient determined by the collision events incurred during the scenario. This formulation enables the evaluation to reflect both task-level behavioral competence and safety performance in a clear and interpretable manner.

BaseScore

BaseScore is composed of scenario-specific task sub-items, with a value range of 0–100, which directly reflects the completion quality of scenario-specific tasks.To improve interpretability, we explicitly decompose the BaseScore of each scenario category into several scenario-specific behavioral sub-items. These sub-items reflect the intended driving objectives in that scenario, such as timely braking, safe bypass, yielding, or post-conflict recovery. The sum of these sub-items forms the final BaseScore for that scenario category.

Table 1: BaseScore composition for the 12 scenario categories in Bench2InterActDrive.

Scenario Category	**BaseScore Components**	**BaseScore**
Lead Vehicle Disappearance	Timely brake and deceleration for static obstacle (50); safely bypass the disappeared front vehicle (30); successfully reach the end point (20)	100
High-Speed Temporary Construction Barrier	Slow down in advance for construction barrier (55); safely detour around the barrier without collision (20); reach the goal after passing the barrier (25)	100
High-Speed Reckless Lane Cutting	Brake response to sudden cut-in vehicle (60); safely bypass the cutting vehicle and avoid conflict (40)	100
Highway Accident Vehicle	Brake and decelerate upon detecting accident vehicle (55); safely bypass the accident area (30); resume normal driving after passing (15)	100
Construction Lane Closure with Truck	Recognize construction zone and decelerate effectively (80); complete passing the construction section by distance target (20)	100

Scenario Category	**BaseScore Components**	**BaseScore**
Roundabout Merge Conflict	Decelerate in response to roundabout traffic (55); yield to oncoming convoy inside roundabout (20); pass through safely and complete route (25)	100
Ghost Probe (Students Crossing Road)	Decelerate when detecting hidden scooter (25); stop completely for sudden pedestrian (55); resume driving after pedestrian leaves (20)	100
Disabled Vehicle On Road	Brake response to broken-down vehicle (40); safely bypass the disabled vehicle (40); resume route after passing (20)	100
E-Bike and Pedestrian Crossing Obliquely	Decelerate for detected e-bike (25); brake fully for crossing pedestrian (50); resume travel after risk clearance (25)	100
Reverse Vehicle Intrusion	Brake in response to reversing vehicle (55); safely bypass the reversing vehicle (20); resume normal route afterward (25)	100
Crazy Motorbike Crossing	Decelerate for erratic motorbike (55); avoid collision with the motorbike (20); resume stable driving after conflict (25)	100
Left Turn with Blind Spot Hidden Car	Brake response to hidden vehicle in blind spot (70); safely pass the intersection and avoid collision (30)	100

Penalty

Penalty_i denotes the collision-based penalty coefficient for scenario i. It is defined according to the collision events incurred by the ego vehicle during the scenario:

- for each collision with a pedestrian, the penalty is multiplied by 0.5;
- for each collision with another vehicle, the penalty is multiplied by 0.6;
- for each collision with a static obstacle, the penalty is multiplied by 0.65;
- if no collision occurs, the penalty remains 1.0.

Formally, let $N_{\text{ped},i}$, $N_{\text{veh},i}$, and $N_{\text{sta},i}$ denote the numbers of collisions with pedestrians, vehicles, and static obstacles, respectively, in scenario i. Then the penalty term is computed as

$$\text{Penalty}_i = (0.5)^{N_{\text{ped},i}}(0.6)^{N_{\text{veh},i}}(0.65)^{N_{\text{sta},i}}.$$

5.3 Average Bench2InterActDrive Score

The overall Bench2InterActDrive score is computed as the average over all N_{B2IAD} extreme interactive scenarios:

$$\overline{S}^{\text{B2IAD}} = \frac{1}{N_{\text{B2IAD}}} \sum_{i=1}^{N_{\text{B2IAD}}} S_i^{\text{B2IAD}}$$

This average reflects a model's ability to handle diverse safety-critical and long-tail scenarios in a behaviorally appropriate way.

5.4 Final Weighted Score

To jointly evaluate *general driving competence* and *extreme-scenario handling ability*, the final leaderboard score is defined as a weighted sum of the Bench2Drive score and the B2IAD score:

$$S^{\text{Final}} = w_{\text{B2D}}\,\overline{S}^{\text{B2D}} + w_{\text{B2IAD}}\,\overline{S}^{\text{B2IAD}}$$

subject to

$$w_{\text{B2D}} + w_{\text{B2IAD}} = 1, \qquad w_{\text{B2D}}, w_{\text{B2IAD}} \geq 0$$

Here, w_{B2D} and w_{B2IAD} are organizer-defined weights released together with the official evaluation toolkit. In practice:

- the Bench2Drive term prevents models from achieving strong final performance through stop-only or overly conservative policies that fail to drive normally;
- the B2IAD term ensures that strong route-following ability alone is not sufficient unless the model also handles safety-critical interactions properly.

Therefore, a competitive submission must perform well on both standard closed-loop routes and extreme interactive scenarios.

5.5 Final Ranking

The final ranking of submissions will be determined according to S^{Final}. Since the benchmark includes both standard driving routes and extreme corner-case scenarios, the resulting leaderboard is intended to reflect a model's overall robustness, safety, and generalization ability, rather than overfitting to a single type of environment or relying on a single behavioral strategy.

6 Submission and Evaluation

Participants who wish to join the 2026 Benchmark Challenge need to complete registration according to the conference or challenge announcement and obtain the official benchmark package from the organizers. The competition is structured as a standardized submission-and-evaluation process: the organizers provide the official benchmark environment and rules, while participants develop their own methods under the provided interface and submit runnable packages for unified evaluation.

After receiving the benchmark toolkit, participants are expected to adapt their end-to-end driving model to the official interface and conduct local validation using the released examples and scripts. The submitted package should contain all necessary files required to run the driving agent in the organizer's evaluation environment. To ensure reproducibility and fairness, the submission should be self-contained, clearly structured, and compatible with the specified software dependencies.

A standard submission is expected to include the following items:

1. the executable agent code adapted to the official benchmark interface;
2. required model weights or checkpoints;
3. configuration files and dependency instructions;
4. a brief method description document explaining the core idea of the submitted approach;
5. any additional scripts needed for loading the model and running inference under the benchmark environment.

During official evaluation, the organizers will run each valid submission on the benchmark platform under the same predefined conditions. The evaluation process will be fully automatic, and all submissions will be tested on the full set of benchmark scenarios. Participants are not allowed to manually modify route definitions, scenario triggers, or runtime conditions during official testing.

The submitted methods will be ranked according to the official benchmark metrics and scoring rules. Depending on the challenge arrangement, rankings may be presented in terms of overall score as well as selected sub-dimensions such as safety, completion, or scenario-specific performance.

In addition to algorithm submission, participants may also be asked to submit a technical report or paper describing their method. This helps promote academic exchange and provides insight into the design choices behind different benchmark results. The final awards, leaderboard announcement, and challenge presentation arrangements will follow the official conference schedule.