# MAPLE: Latent Multi-Agent Play for End-to-End Autonomous Driving

Rajeev Yasarla[1] · Deepti Hegde[1*] · Hsin-Pai Cheng[1*] · Shizhong Han[1*] · Yunxiao Shi[1] · Meysam Sadeghigooghari[2]
Hanno Ackermann[1] · Litian Liu[1] · Pranav Desai[2] · Fatih Porikli[1] · Mohammad Ghavamzadeh[1] · Hong Cai[1]
[1] Qualcomm AI Research [2] Qualcomm Technologies, Inc. * Equal contribution

NeurIPS 2026

## Motivation

### How would other drivers respond to our actions?

Logs capture one realized future. A new ego decision can change whether nearby drivers yield, brake, or merge. Fixed playback misses these reactive responses. Small deviations then compound in driving.

### Why generative rollouts matter

Latent rollouts explore alternative reactions and expose the ego to evolving interactions. This creates richer feedback without rendering future images or using an external simulator.

## Contributions & Novelty

### 1 Reactive latent rollouts

The VLA jointly evolves ego and nearby-agent tokens over multiple steps. Learned policies control reactive agents; a motion head handles background traffic. Plans update the next scene.

### 2 Diversity-aware RL

Shared safety and progress goals can lead to similar policies. Diversity rewards separate behavior descriptors, exposing the ego to varied responses alongside safety and progress feedback.

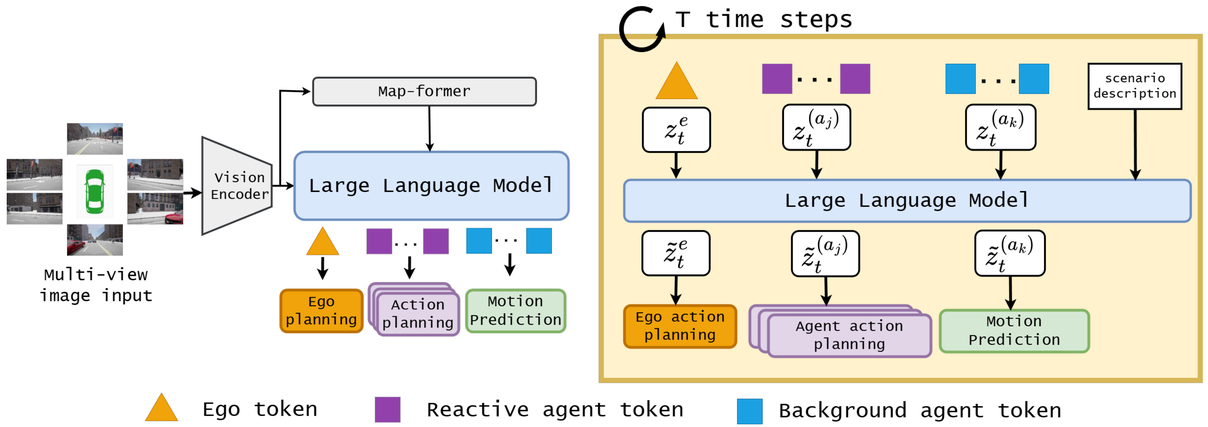
### 3 Learn, then interact

Rollout imitation grounds planning in driving logs; GRPO refines multi-agent interactions.
The rollout loop is disabled at deployment.

## Fail2Drive Generalization

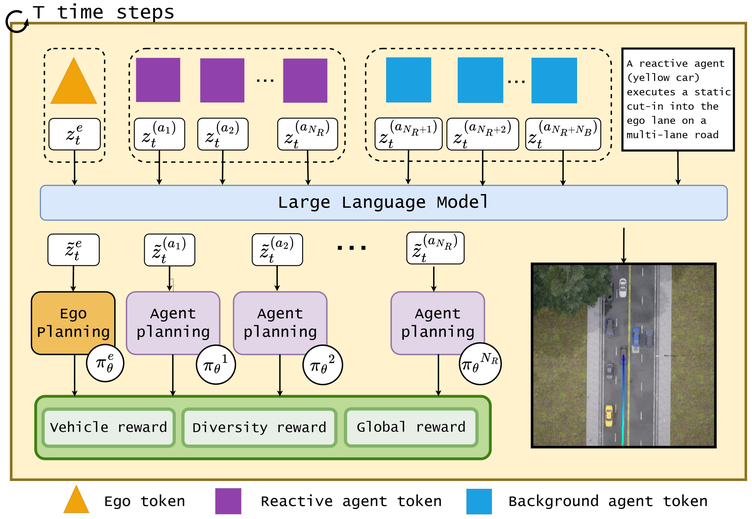
| | In-distribution | | Out-of-distribution | |
|---|---|---|---|---|
| **Method** | **DS ↑** | **SR ↑** | **DS ↑** | **SR ↑** |
| SimLingo | 80.6 | 79.3 | 71.7 | 55.0 |
| HiP-AD | 74.1 | 70.7 | 67.1 | 56.7 |
| **MAPLE** | **85.3** | **82.1** | **76.1** | **63.7** |
| **MAPLE*** | **86.7** | **82.9** | **79.5** | **65.8** |

## Overview: Reactive Latent Rollout




Joint ego and reactive-agent planning over T rollout steps.

## Multi-Agent Play




### Global Reward

Collision-free rollouts and scene-level safety feedback.

### Vehicle Reward

Progress, collision avoidance, and time-to-collision signals.

### Diversity Reward

Distinct behaviors across policies broaden the ego's experience of reactive traffic interactions.

## Closed-Loop Comparison

Bench2Drive · 220 test routes

| Method | DS ↑ | SR ↑ | MA ↑ |
|---|---|---|---|
| Qwen2.5 | 63.9 | 31.6 | 25.6 |
| ORION | 77.7 | 54.6 | 54.7 |
| GeRo | 81.9 | 60.1 | 66.5 |
| SimLingo (TP) | 85.1 | 67.3 | 67.0 |
| HiP-AD* | 86.8 | 69.1 | 65.9 |
| **MAPLE** | **85.2** | **67.1** | **70.6** |
| **MAPLE*** | **88.3** | **70.5** | **73.2** |

DS: driving score; SR: success rate (%).
MA: mean multi-ability (%); all higher better.
TP: target point; others: nav. command.

*: geometric paths + SimLingo controller.
Training and supervision differ.

## Ablation: Training & Play

| Configuration | DS ↑ |
|---|---|
| **Training curriculum (ego-only)** | |
| Baseline | 68.1 |
| + state estimation pretraining | 70.4 |
| + rollout SFT | 77.5 |
| + ego-only rollout RL | 80.4 |
| **Multi-agent design (after rollout RL)** | |
| + reactive agents, shared policy | 82.1 |
| + distinct reactive-agent policies | 83.4 |
| **+ diversity reward (full MAPLE)** | **85.2** |

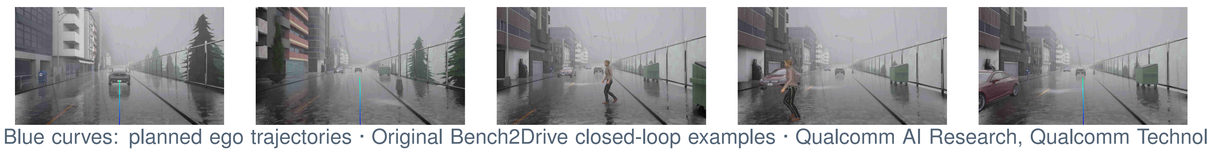
**Diversity: +1.8 driving-score points**

## Latent Rollout Realism

| Method | Meta ↑ | Kin. ↑ | Int. ↑ |
|---|---|---|---|
| DeepSight | 0.490 | 0.413 | 0.608 |
| GeRo | 0.553 | 0.454 | 0.620 |
| **MAPLE** | **0.572** | **0.476** | **0.703** |

Adapted Waymo trajectory metrics

## Qualitative: Pedestrian Crossing



## Qualitative: Vehicles and Cyclists



Blue curves: planned ego trajectories · Original Bench2Drive closed-loop examples · Qualcomm AI Research, Qualcomm Technologies, Inc.