JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

CVPR 2023 Workshop on Autonomous Driving
https://cvpr2023.wad.vision/

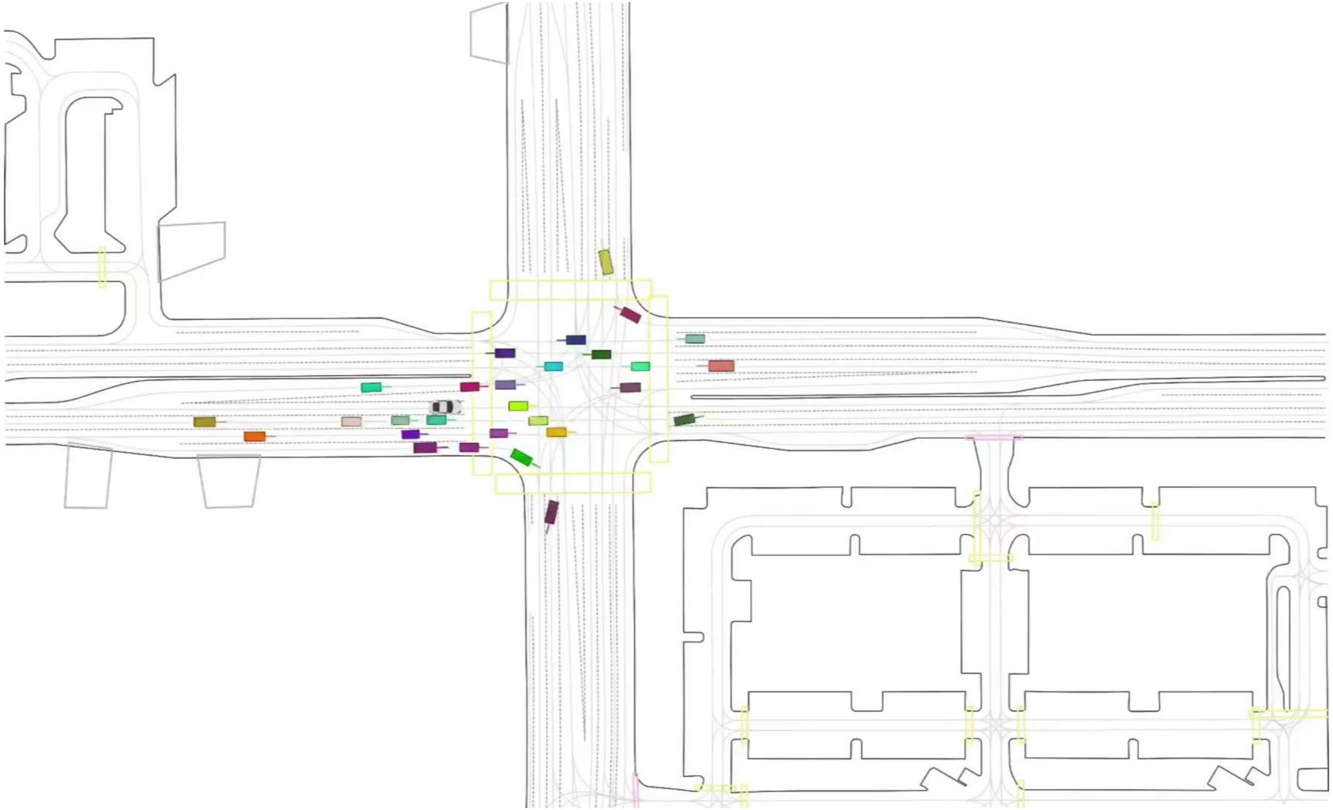
# MULTIVERSE TRANSFORMER

## 1st Place Solution for Waymo Open Sim Agents Challenge 2023

**Yu Wang, Tiebiao Zhao, Fan Yi**
**Pegasus**

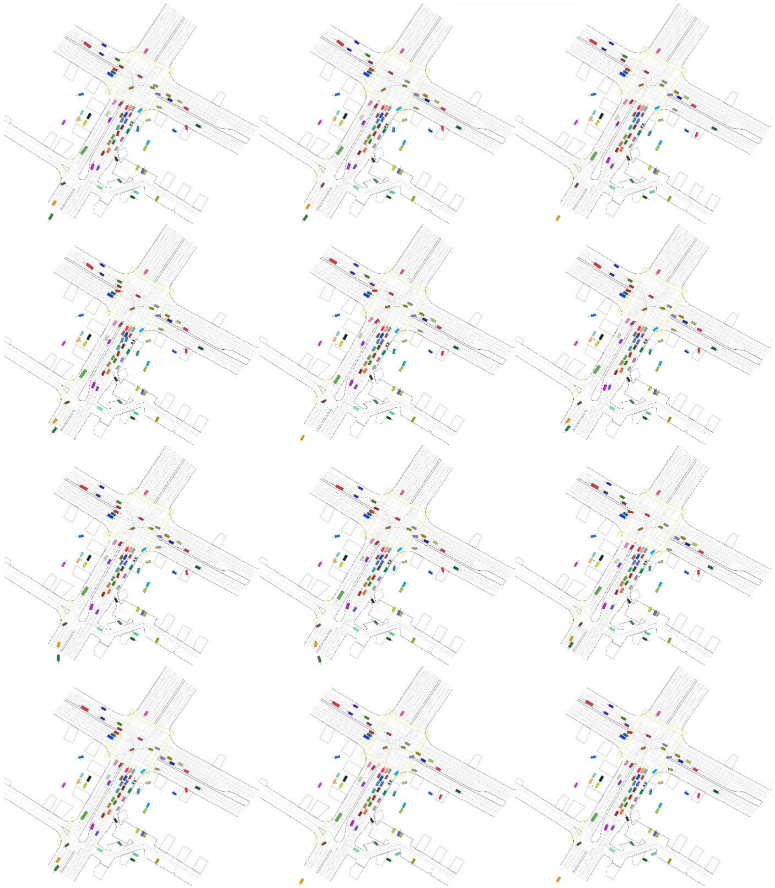
# MultiVerse

Every decision a driver makes, at every fleeting moment, spawns a multitude of parallel universes.



# MultiVerse

Our multiverse simulator generates a multitude of parallel universes. Each one mirrors the logged real-world driving data, yet diverges subtly in its own unique way.



# Overview

- Problem Formulation
- Our Solution:
  - MultiVerse Transformer for Agent (MVTA) Simulation
- Experimental Results
  - WOSAC Leaderboard
  - Qualitative Results
- Summary and Future works

# Problem Formulation

- Waymo Open Sim Agents Challenge (WOSAC)
  - Given scene context, including map and past positions of the agents (both world agents and ADV), simulate states of the agents at 0.1s intervals for the upcoming 80 timesteps

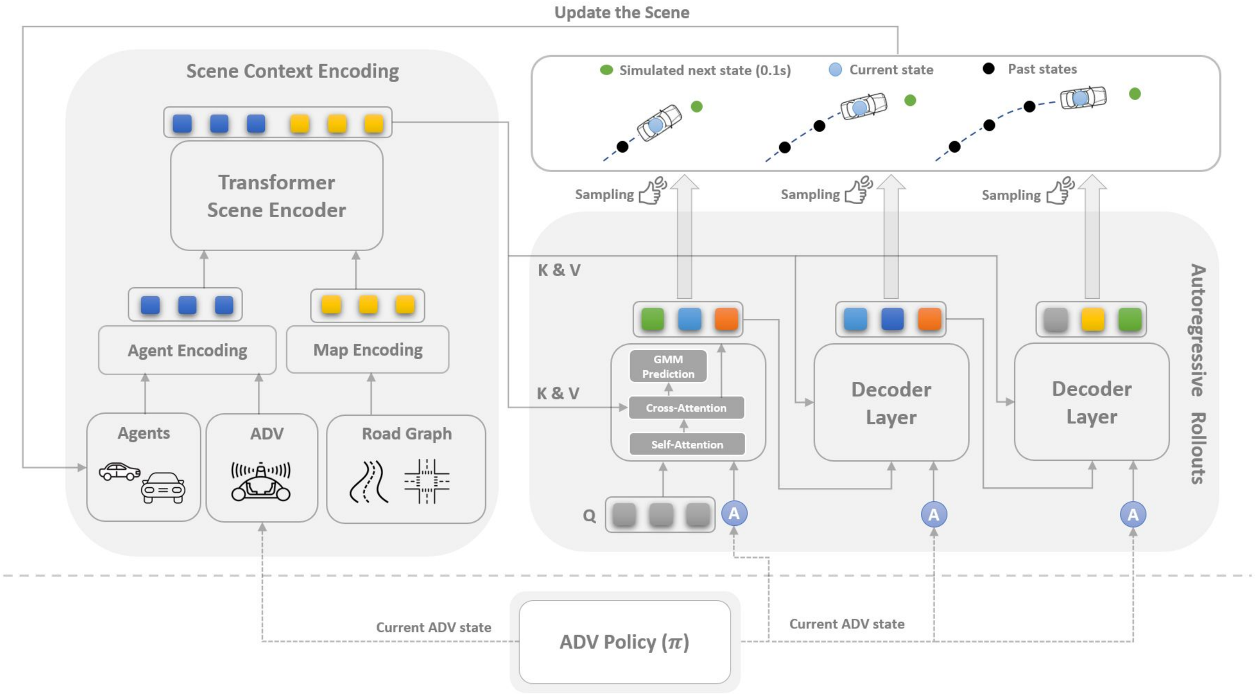
- The constraints
  - Simulator must be closed-loop, and run in autoregressive manner
  - World agents and ADV must be conditionally independent, i.e., ADV component can be replaced with any arbitrary policy or planner

# Traffic Agents Simulation

- Motion Prediction
  - **Agent-centric** v.s. Scene-centric
  - **Autoregressive** v.s. Non-autoregressive
- Traffic Simulation
  - **Learning-based generative model** v.s. Heuristic-based model encoding traffic rules
  - **Closed loop** v.s. open loop
- Our work is inspired by **TrafficSim** and **MTR**

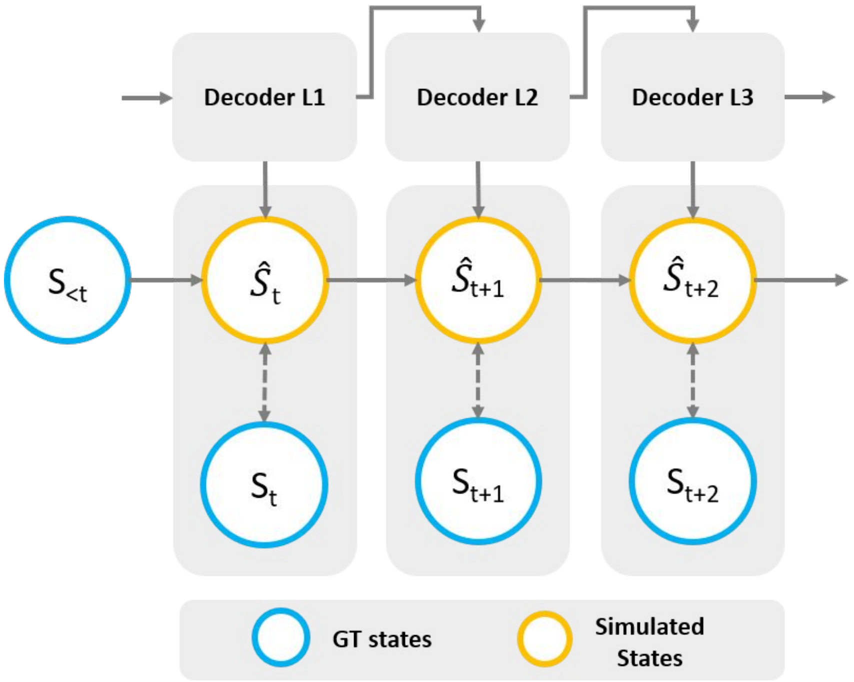
# MultiVerse Transformer

- **M**ulti**V**erse **T**ransformer for **A**gent (**MVTA**) simulation: encoder-decoder transformer with autoregressive rollout for closed-loop simulation



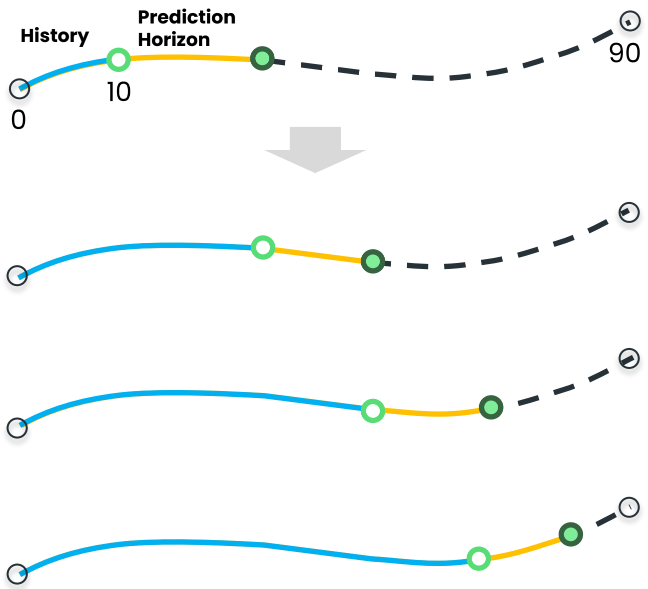
# Training

- End-to-end training, with loss calculated at each timestep



# Training

- **Training sample generation**: randomly pick a point to separate the trajectory to history and future components




- **Training Loss**

**Negative log-likelihood loss**: maximizing the likelihood of ground-truth trajectory

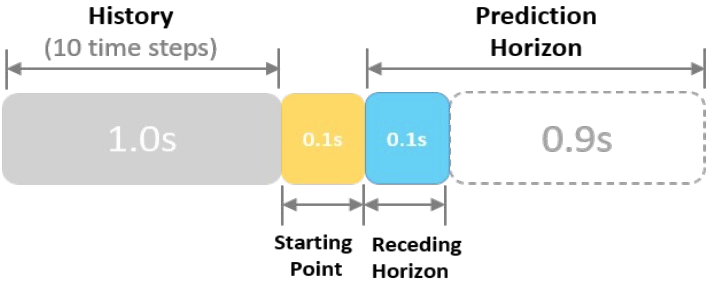
$$\mathcal{L}_{NLL} = -\log\mathcal{N}(S_x - \mu_x, \sigma_x; S_y - \mu_y, \sigma_y; \rho)$$

**Total loss**: NLL with L1 losses for velocity and heading

$$\mathcal{L}_{total} = \sum_t \lambda_1 \mathcal{L}^t_{NLL} + \lambda_2 \mathcal{L}^t_{Vel} + \lambda_3 \mathcal{L}^t_{\theta}$$

# Receding Prediction Horizon

- Receding prediction horizon: prediction horizon is 1s, but only the waypoint of the initial 0.1s is utilized, with the remaining prediction being discarded.



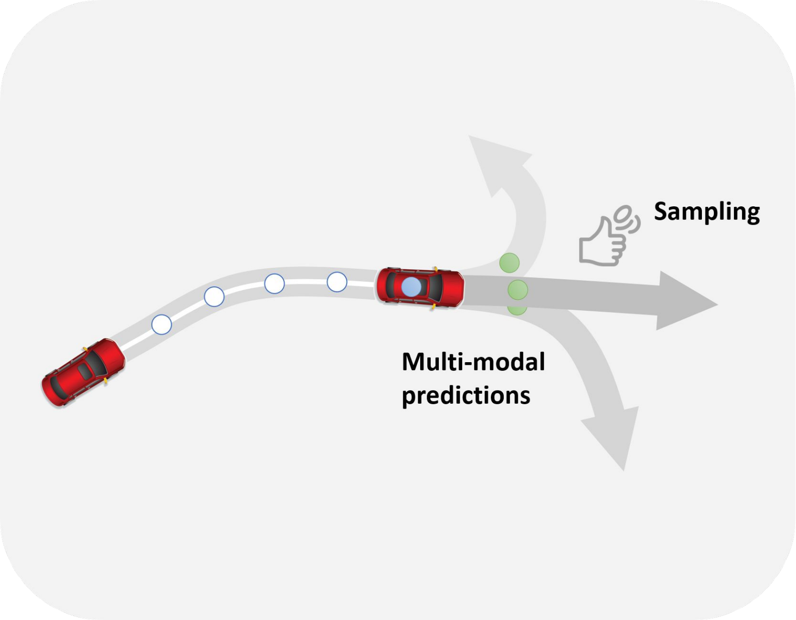

➕ Promoting multi-modal diversity

➕ Reducing compounding error during autoregressive execution

➕ More flexibility in the inference setup

# Inference

- Top-K sampling: applied at periodic intervals to strike a balance between realism and diversity

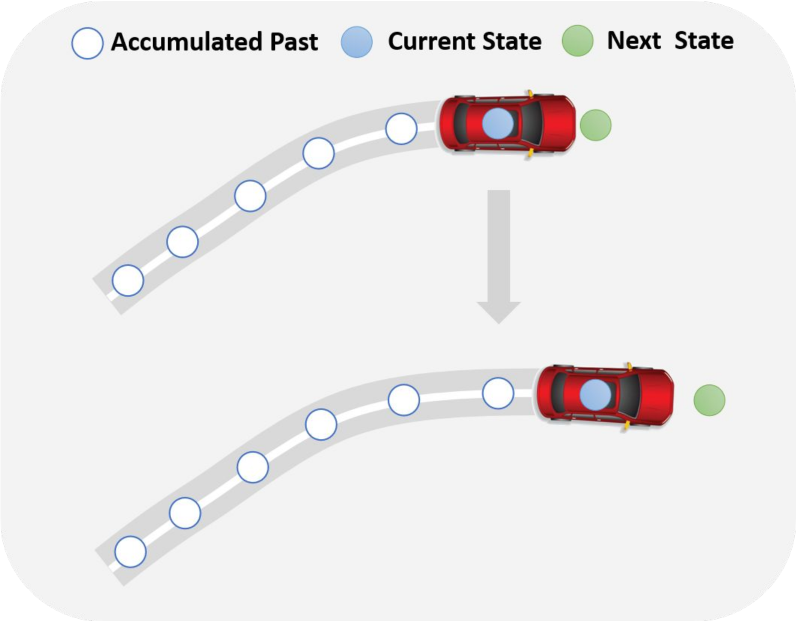



➕ Promoting multi-modal diversity, as opposed to selecting the trajectory with the highest likelihood

➖ Susceptible to the compounding error and could generate trajectories with unrealistic kinematic motions or even drift

# Inference

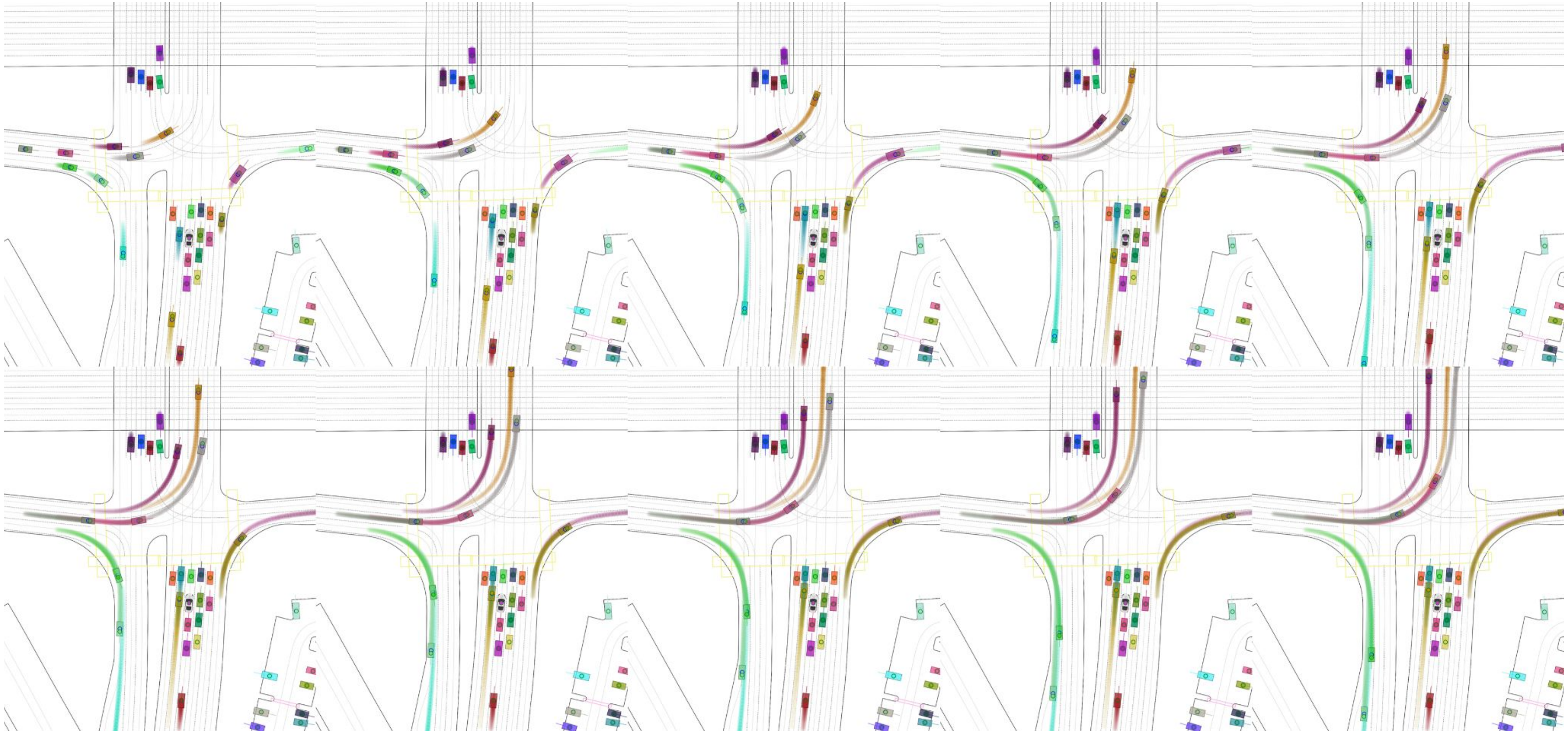
- Variable-length history: aggregate the past trajectory overtime




➕ Aligned with training which also uses variable-length history

➕ Longer history enhances the stability of the simulation by reducing the potential for compounding error

# Autoregressive Rollout



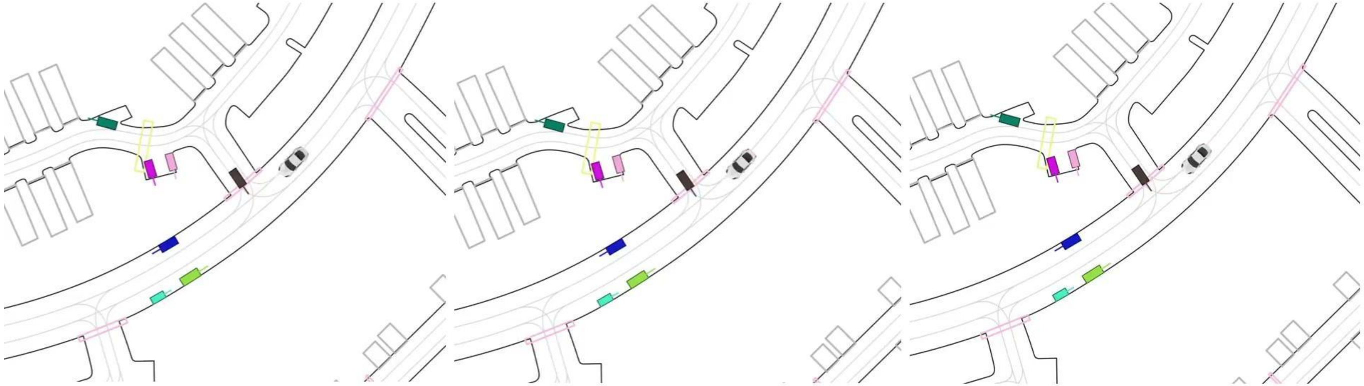
# WOSAC Leaderboard

- WOSAC Leaderboard
  - 486,995, 44,097, and 44,920 scenarios in the training, validation, and test set, respectively
  - Main metric is the realism meta-metric, aggregating a group of component metrics (kinematic, interactive, and map-based metrics)
  - **MTVE**: enhanced version of **MVTA** with three model variants

| WAYMO | META METRIC | KINEMATIC | | | | INTERACTIVE | | | MAP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **LEADERBOARD** | **REALISM** | LINEAR SPEED | LINEAR ACCEL. | ANG. SPEED | ANG. ACCEL. | DIST. TO OBJ. | COLLISION | TTC | DIST. TO ROAD | OFFROAD | minADE (↓) |
| **MVTE** (ours) | **0.5168** | 0.4426 | 0.2218 | **0.5353** | **0.481** | **0.382** | 0.4509 | **0.832** | **0.6641** | **0.6409** | 1.677 |
| **MVTA** (ours) | 0.5091 | 0.4365 | 0.22 | 0.533 | 0.4805 | 0.3729 | 0.4359 | 0.8298 | 0.6545 | 0.6288 | 1.8698 |
| MTR+++ | 0.4697 | 0.4119 | 0.1066 | 0.4838 | 0.4365 | 0.3457 | 0.4144 | 0.7969 | 0.6545 | 0.577 | 1.6817 |
| CAD | 0.4321 | 0.3464 | **0.2526** | 0.4327 | 0.311 | 0.33 | 0.3114 | 0.7893 | 0.6376 | 0.5397 | 2.3146 |
| multipath | 0.424 | 0.4318 | 0.2304 | 0.0193 | 0.0355 | 0.3493 | **0.4854** | 0.8111 | 0.6372 | 0.613 | 2.0517 |
| sim_agents_tutorial | 0.3941 | 0.3143 | 0.1738 | 0.4785 | 0.4631 | 0.2641 | 0.2671 | 0.7709 | 0.5575 | 0.4111 | 3.6198 |
| QCNeXt | 0.392 | **0.4773** | 0.2424 | 0.3252 | 0.1987 | 0.3759 | 0.3244 | 0.7569 | 0.6099 | 0.36 | **1.083** |
| sim_agents_tutorial | 0.3201 | 0.3826 | 0.0999 | 0.0318 | 0.0391 | 0.2909 | 0.336 | 0.7549 | 0.5251 | 0.3804 | 3.108 |
| linear_extrapolation_baseline_tutorial | 0.2576 | 0.0745 | 0.1659 | 0.0187 | 0.0348 | 0.2221 | 0.2211 | 0.7551 | 0.479 | 0.3352 | 7.5148 |

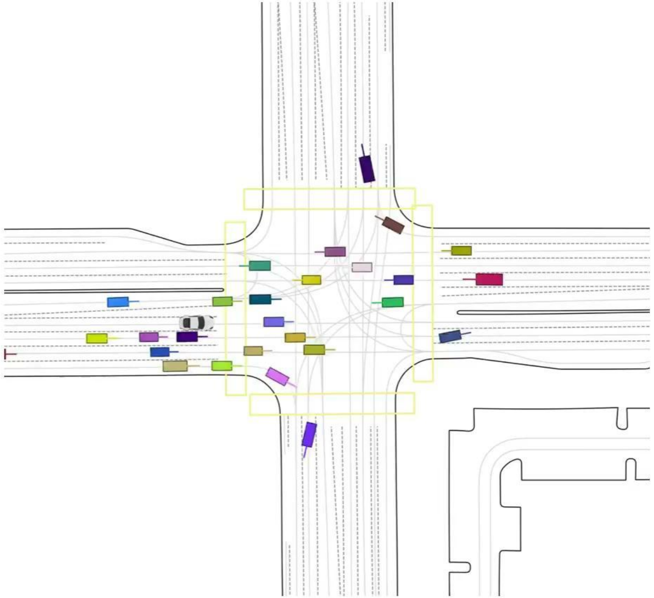
# Qualitative Result

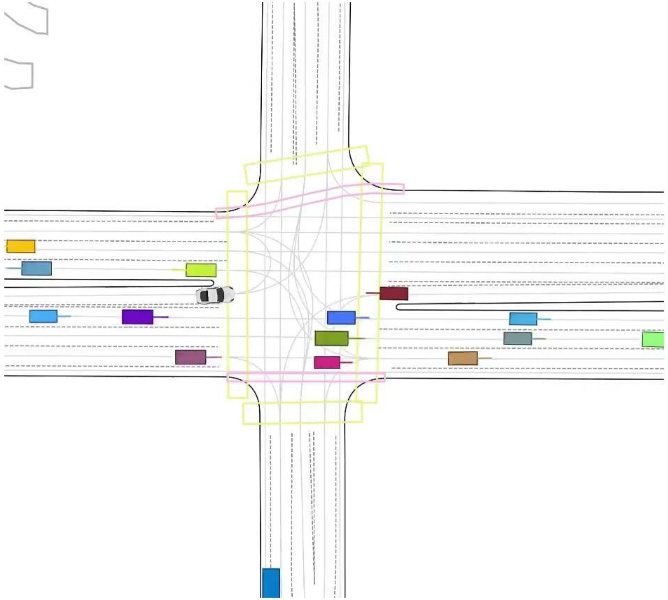
A vehicle waiting to get onto the main road, the car turns left or right, or keeps waiting. ADV also demonstrates multi-modal behavior.



# Qualitative Result

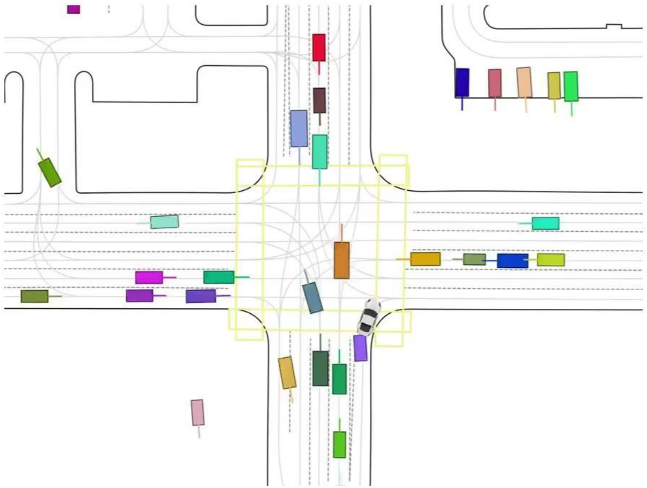
Vehicles paused at an intersection, ready to make an unprotected left turn, waiting for the oncoming traffic to clear.

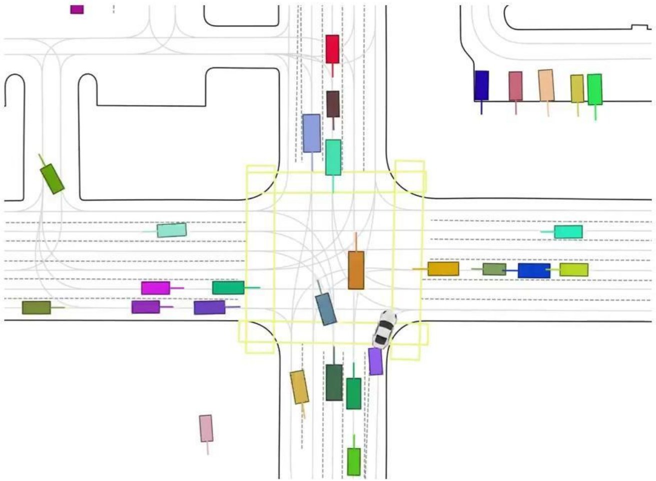




# Qualitative Result

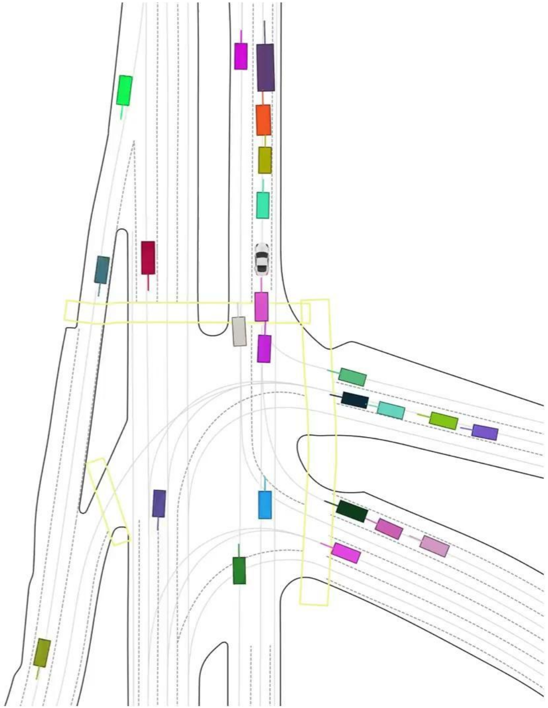
ADV makes a slow right-turn, forcing the agent behind it to slow down or stop.





# Qualitative Result

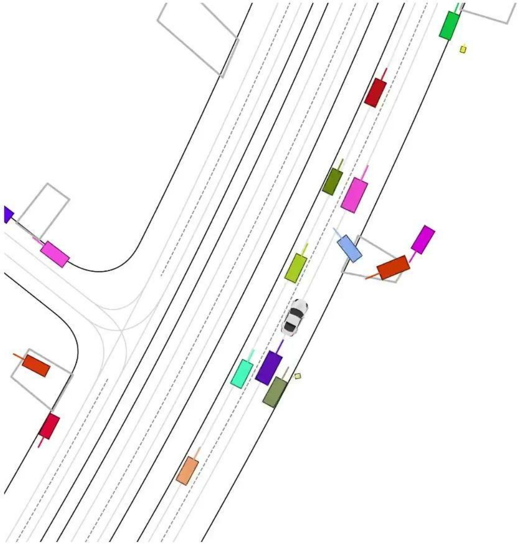
A congested right lane, blocked by a large truck, and a free-flowing left lane. Trapped cars attempt to switch lanes, overtaking the ADV and the remaining slow traffic.

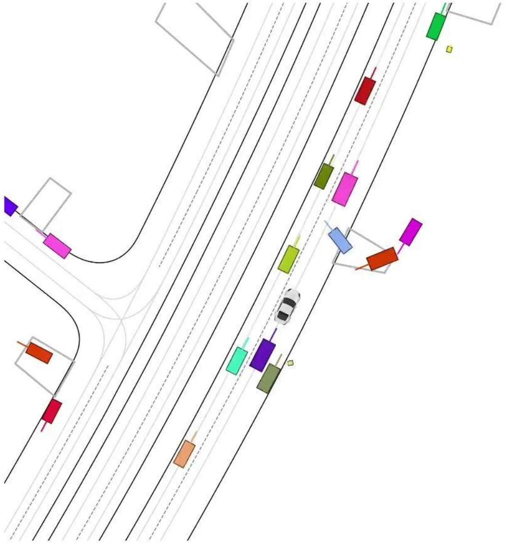


# Qualitative Result

ADV slows down to allow a car merging onto the main road from a driveway. The car behind ADV overtakes it.

In the second simulation, ADV proceeds straight instead of waiting.





# Summary and Future Works

- Transformer-based generative model for closed-loop traffic agents simulation
- Novel training, sampling and inference strategies promoting a high degree of realism and diversity
- First place on WOSAC 2023
- Future works
  - Collision avoidance loss
  - Scene-centric and Diffusion-based simulation approaches