ETH zürich

Institute of Dynamic Systems and Control



Rule-Constrained Adversial Scenario Generation

Presenter: Yash Purswani
RobotX 2026 Fellow

Professor Frazzoli's Group
IDSC, D-MAVT, ETH Zürich

Motivation and Research Challenge

1 · Background

AVs Scaling up
Increasing number of Robotaxis on real roads

Hidden Risk
Rare crash cases almost never appear

Costly Testing
Road trials are expensive and time-taking

2 · Why Adversarial Scenarios?

Regulators
Auditable certification

AV Manufactures
Identify failures early

The Public
Safety and less anxiety

OBJECTIVES

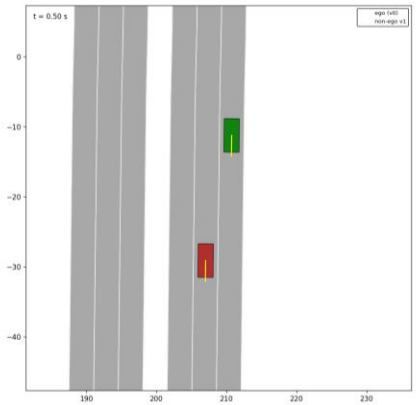
Maximize ego rule violations score

Satisfy all rules for the non-ego

Challenge How can we generate simulation scenarios that are both **critical** and **credible** to predict real-world performance?

State of the Art Techniques

LLM-Based – ChatScene

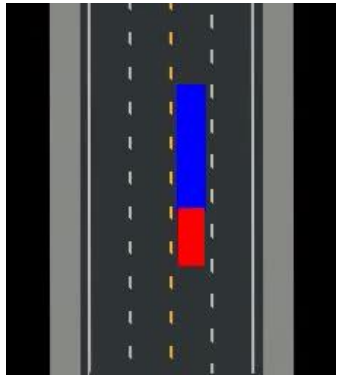


✓ Natural-language instruction

✗ Limit to pre-defined non-ego behaviors

(Zhang et.al CVPR24)

Generative Models – DiffScence

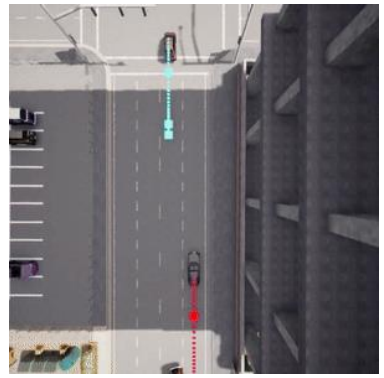


✓ Human-like non-ego behaviour

✗ Can be overly aggressive

(Xu et.al AAAI'25)

RL-Based – FREA



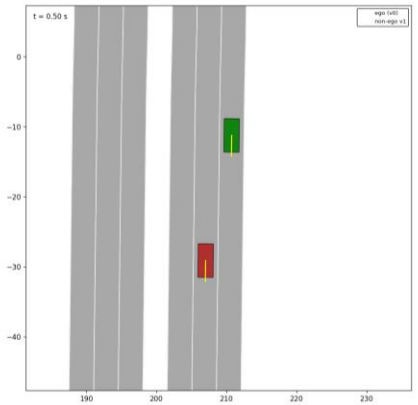
✓ Flexible objective and can generate specified scenarios

✗ Optimizes for ego feasibility and no constraints in non-ego behaviour

(Chen et.al CoRL'22)

State of the Art Techniques

LLM-Based – ChatScene

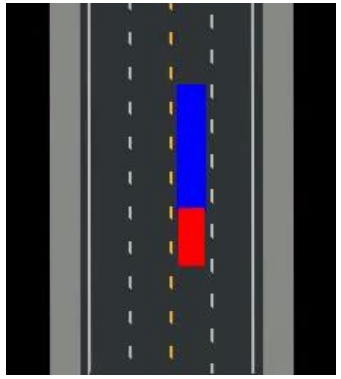


✓ Natural-language instruction

✗ Limit to pre-defined non-ego behaviors

(Zhang et.al CVPR24)

Generative Models – DiffScence

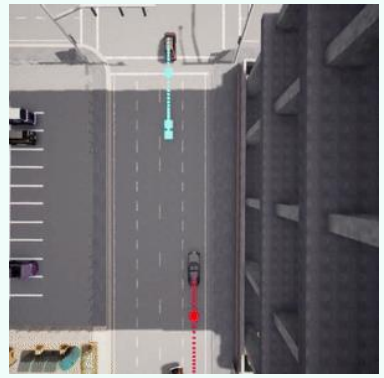


✓ Human-like non-ego behaviour

✗ Can be overly aggressive

(Xu et.al AAAI'25)

RL-Based – FREA



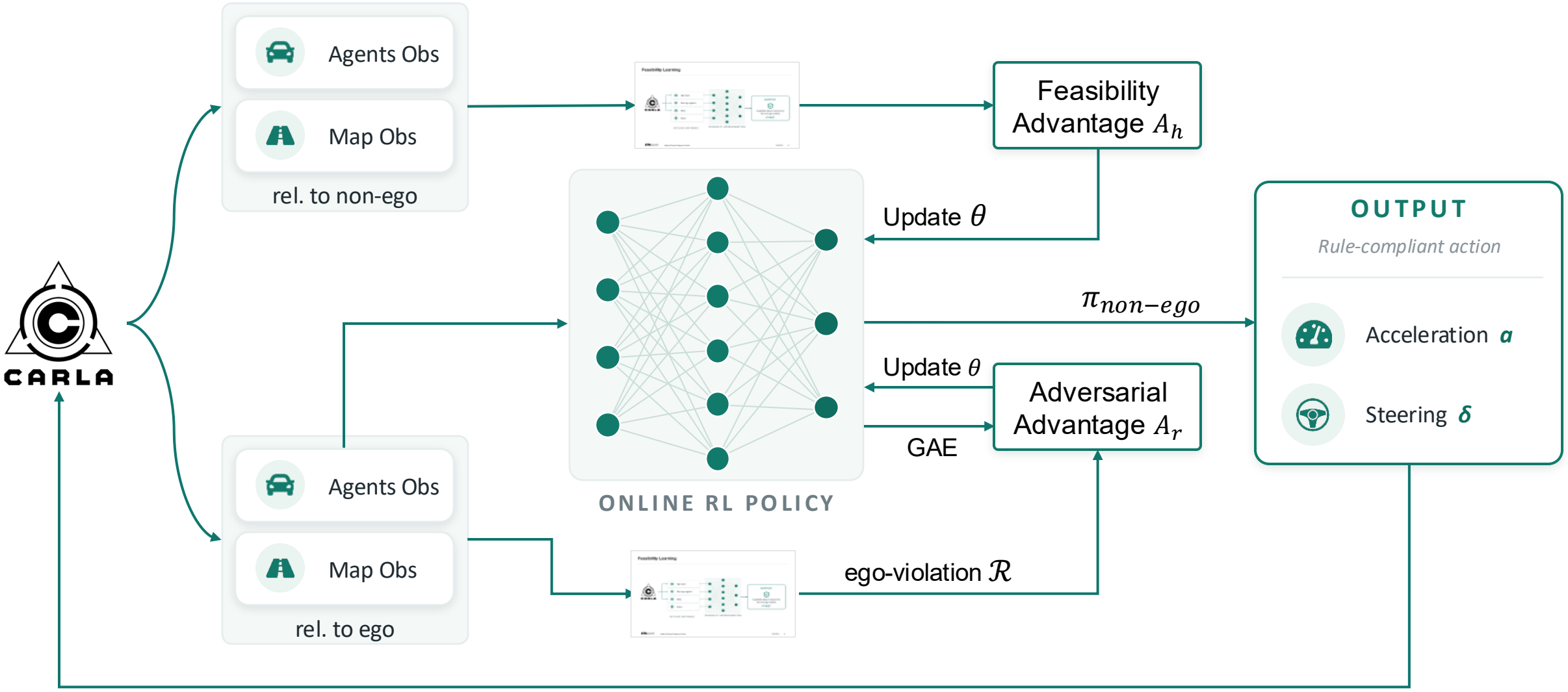
✓ Flexible objective and can generate specified scenarios

✗ Optimizes for ego feasibility and no constraints in non-ego behaviour

(Chen et.al CoRL'24)

OUR AIM: RL-driven adversarial scenario generation with rule-constrained non-ego agents to ensure realistic traffic behaviour

Feasibility-Dependent Adversarial Policy for CBV



Feasibility Learning

Rule-Violation Function

$$h(s^{\mathrm{AV}}) = \begin{cases} -1, & \text{if } r(s^{\mathrm{AV}}) \leq 0, \\ r(s^{\mathrm{AV}}), & \text{if } r(s^{\mathrm{AV}}) > 0. \end{cases}$$

Value Function Definations

$$V_h^*(s^{\mathrm{AV}}) := \min_{\pi^{\mathrm{AV}}} \max_{t \in \mathbb{N}} h\left(s_t^{\mathrm{AV}}\right), \quad s_0^{\mathrm{AV}} = s^{\mathrm{AV}}, \quad a_t^{\mathrm{AV}} \sim \pi^{\mathrm{AV}}(\cdot \mid s_t^{\mathrm{AV}})$$

$$Q_h^*(s^{\mathrm{AV}}, a^{\mathrm{AV}}) := \min_{\pi^{\mathrm{AV}}} \max_{t \in \mathbb{N}} h\left(s_t^{\mathrm{AV}}\right), \quad s_0^{\mathrm{AV}} = s^{\mathrm{AV}}, \quad a_0^{\mathrm{AV}} = a^{\mathrm{AV}}, \quad a_{t+1}^{\mathrm{AV}} \sim \pi^{\mathrm{AV}}(\cdot \mid s_{t+1}^{\mathrm{AV}})$$

If $V_h^*(\mathrm{s}_{\mathrm{AV}}) \leq 0 \Rightarrow \exists\, \pi^{\mathrm{AV}}$, which is optimal and satisfies the constraints.
If $V_h^*(\mathrm{s}_{\mathrm{AV}}) > 0 \Rightarrow \forall\, \pi^{\mathrm{AV}}$, no policy is safe and satisfies the constraints.

Finally , the LFR is defined as follows:

$$S_f^* := \{s^{\mathrm{AV}} | V_h^*(\mathrm{s}_{\mathrm{AV}}) \leq 0\}$$

Feasibility-Dependent Adversarial Policy for CBV

Adversarial RL Formaulation (PPO)

$$L(\theta) = \mathbb{E}_{\pi_{\theta_k}}\left[\min\left(r_t(\theta)\, A^{\pi_{\theta_k}}(s,a),\ \text{clip}\ (r_t(\theta),\, 1-\epsilon,\, 1+\epsilon) A^{\pi_{\theta_k}}(s,a)\right)\right]$$

Advantage Formulation

$$A^{\pi_{\theta_k}}(s,a) = A_r^{\pi_{\theta_k}}(s,a) \cdot I(s,s') + A_h^*(s_{\text{AV}}, a_{\text{AV}}) \cdot \left(1 - I(s,s')\right)$$

A_r is the reward advantage which incentivizes the adversarial ojective.

A_h^* is the feasibilty advantage which keeps the ego vehicle in the safe zone.

$I(s,s')$ is an indiactor function which $= 1$ if $V_h^*(g(s)) \leq 0 \wedge V_h^*(g(s') \leq 0$, i.e. when the AV vehicle is in the feasibility region, otherwise 0.

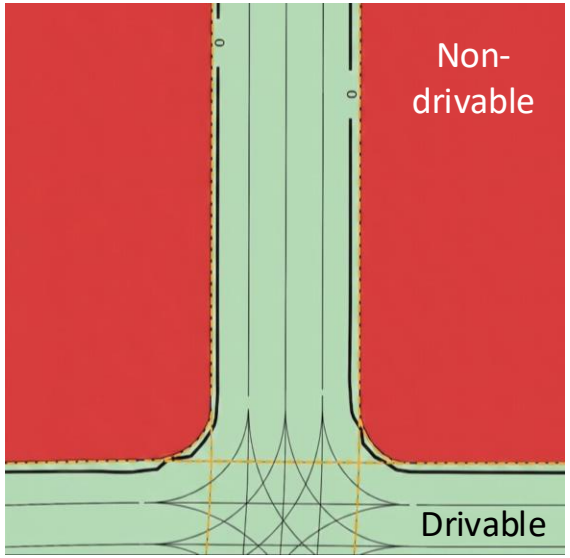
Reward Function

The rewaard distribution used during the experiments is as follows

$$R_t = d(\text{CBV}_{t-1}, \text{Goal}_{t-1}) - d(\text{CBV}_t, \text{Goal}_t) - 50 * r_t^{\text{collison}} + 5 * r(s^{\text{ego}}) + 15 * r_t^{\text{finish}}$$

Rule Definitions

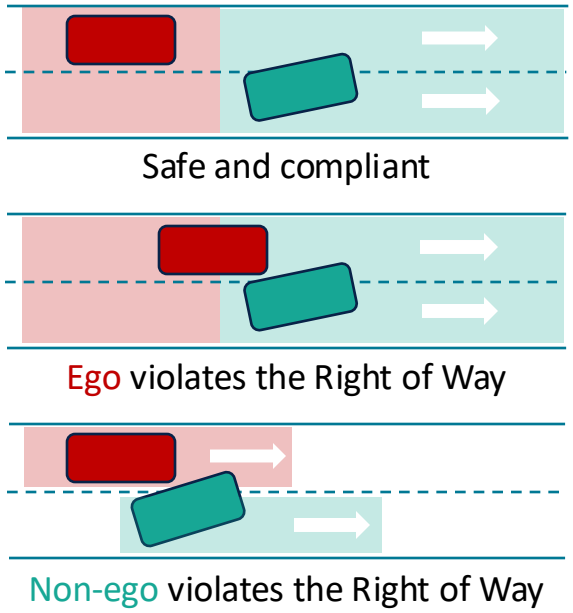
Drivable Region Rule



Max. Speed Rule

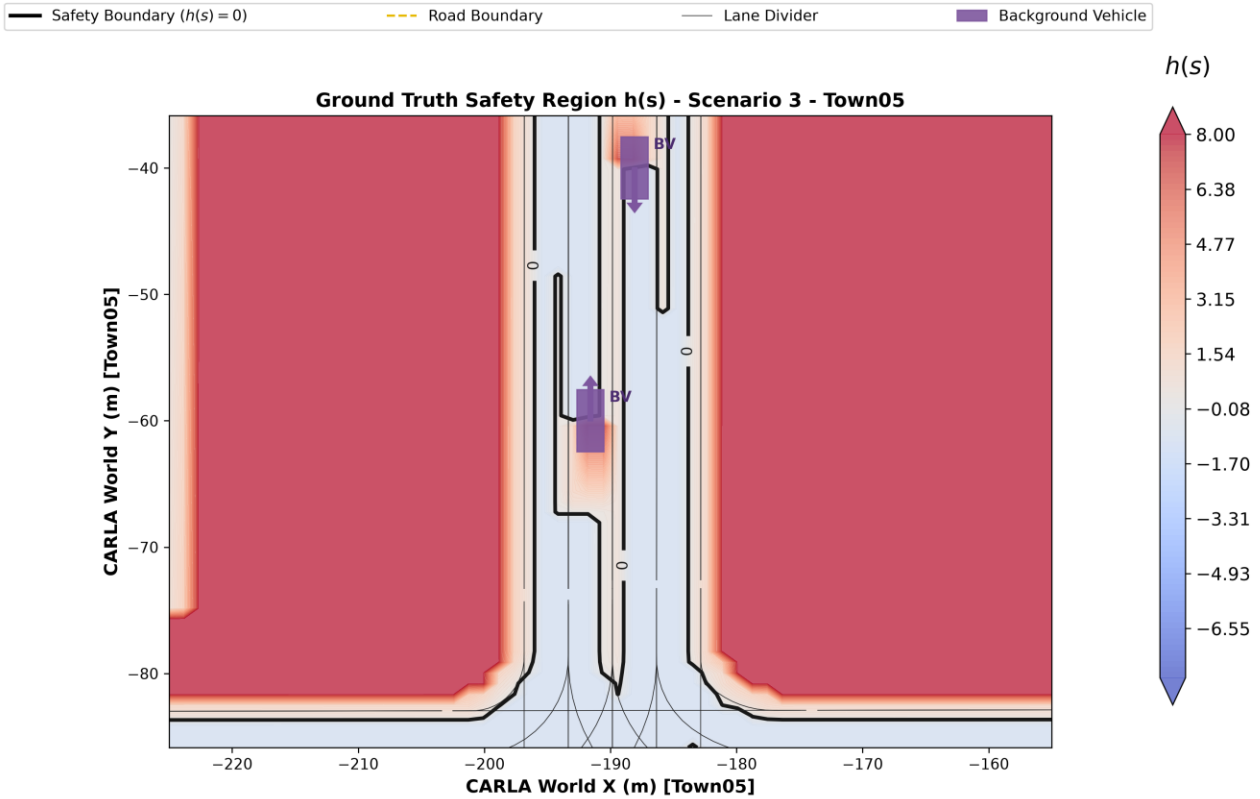


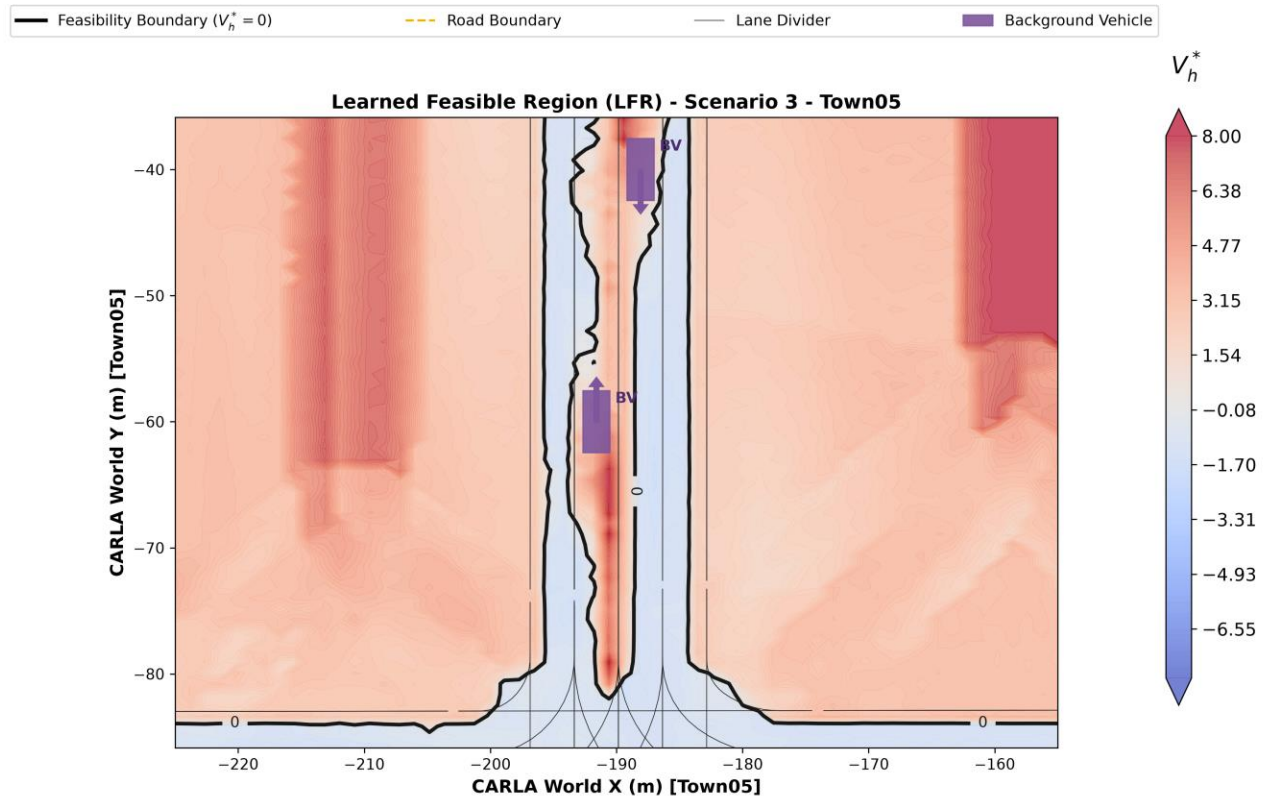
Right-of-Way Rule



The severity of each rule violation is aggregated into a total rule-violation score $r(s^{\mathrm{AV}})$.

Feasibility Learning Results





Experimental Results

Performance Across Different Rule Constraints

Rule Violations	Original FREA		Modified FREA	
	Ego	Non-Ego	Ego	Non-Ego
RightOfWay Rule	0.148 ± 0.025	0.784 ± 0.146	0.138 ± 0.023	0.268 ± 0.053
DrivableRegion Rule	0.000 ± 0.000	0.202 ± 0.046	0.000 ± 0.000	0.034 ± 0.012
MaxSpeed Rule	0.000 ± 0.000	0.292 ± 0.085	0.000 ± 0.000	0.143 ± 0.050
Total Violation	14.844 ± 2.513	**80.721 ± 15.082**	13.800 ± 2.367	**27.309 ± 5.393**

Rule Violation: $\mathrm{rule_violation} \times \mathrm{freq}$ aggregated over 70 episodes
Total Violation: $100 \times r_{\mathrm{RoW}} + 10 \times r_{\mathrm{driv}} + r_{\mathrm{speed}}$

Qualitative Comparison: Baseline vs. Rule-Constrained Policy



Original FREA



Modified FREA

Key Takeaways:

- LFR provides an effective surrogate for evaluating rule violations.
- Feasibility-guided learning significantly reduces non-ego rule violations.
- Ego violations remain policy-dependent, limiting the improvement observed across methods.

ETH zürich

Thank You
Q/A

Yash Purswani
RobotX 2026 Fellow
ypurswani@student.ethz.ch

ETH Zürich
Institute of Dynamic Systems and Control
ML K42
Sonneggstrasse 3
8092 Zürich
Switzerland