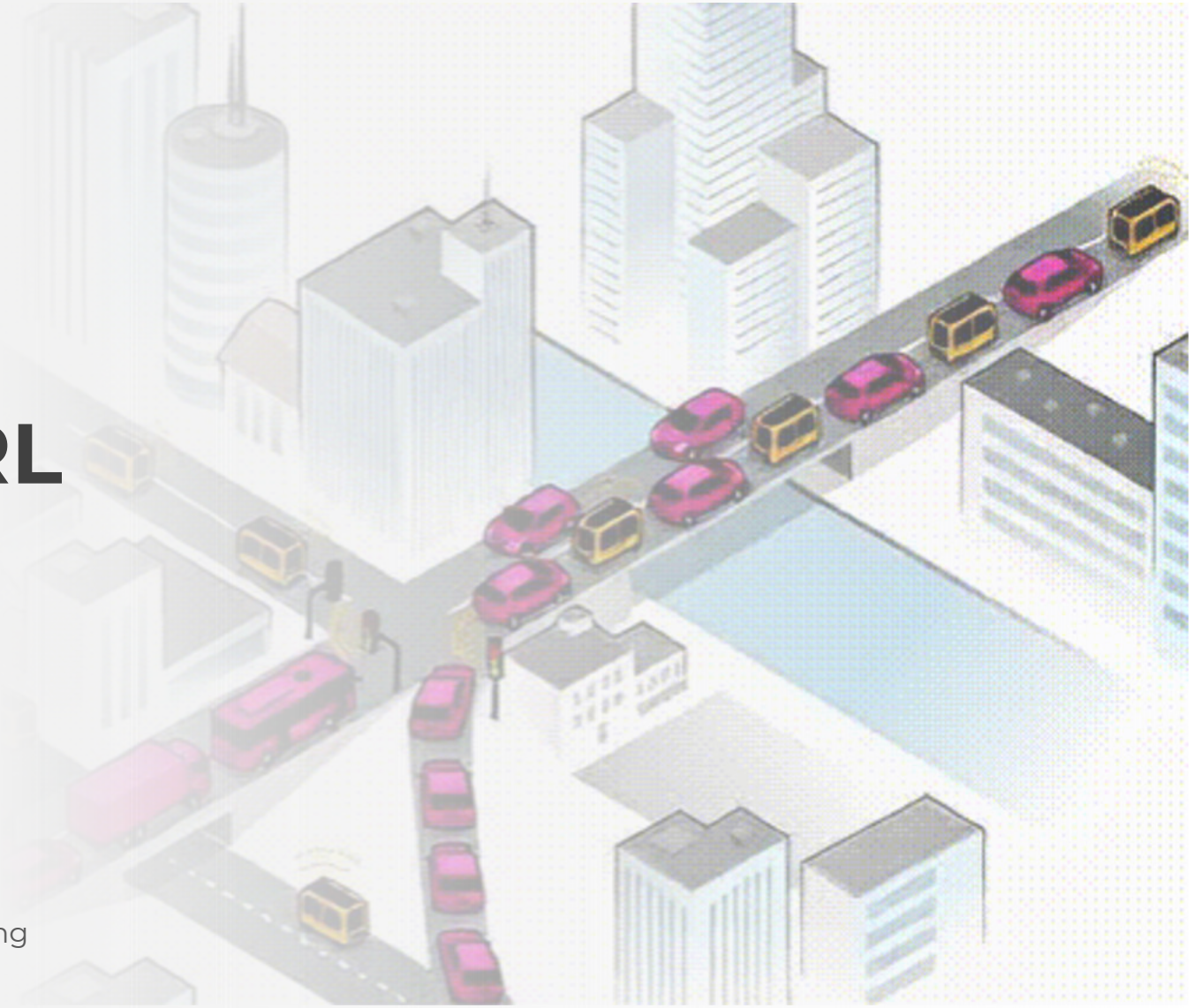
EWRL18
September 17-19 2025, Tübingen

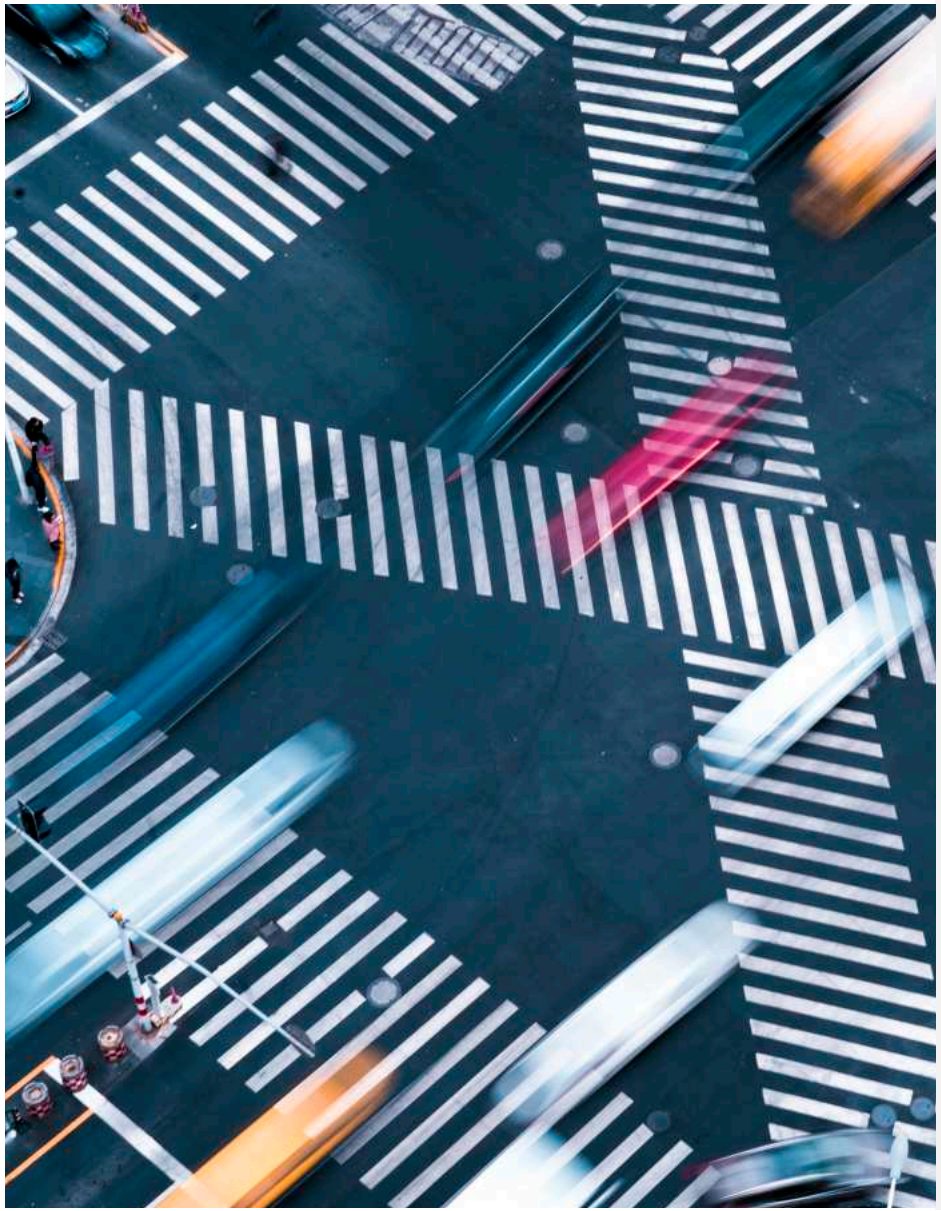
Routing Autonomous Vehicles Using RL

Progress & Future Directions



Onur Akman
Jagiellonian University, Kraków

Tübingen, 17/09/2025
18th European Workshop on Reinforcement Learning

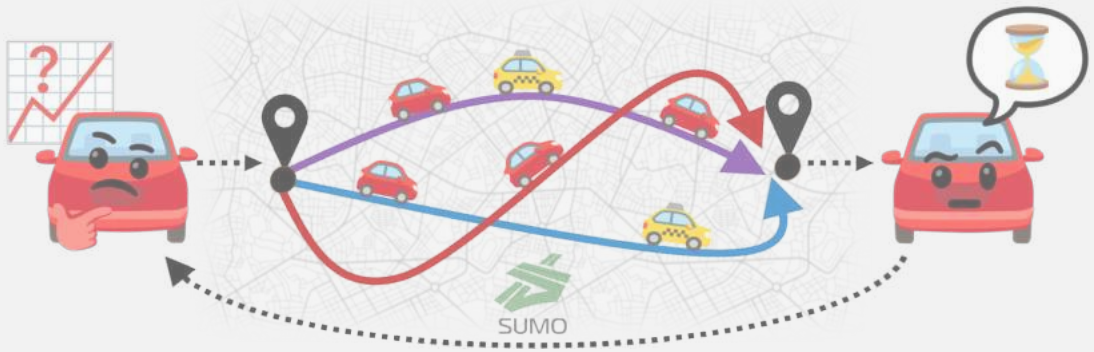


Content

Problem

Scenario

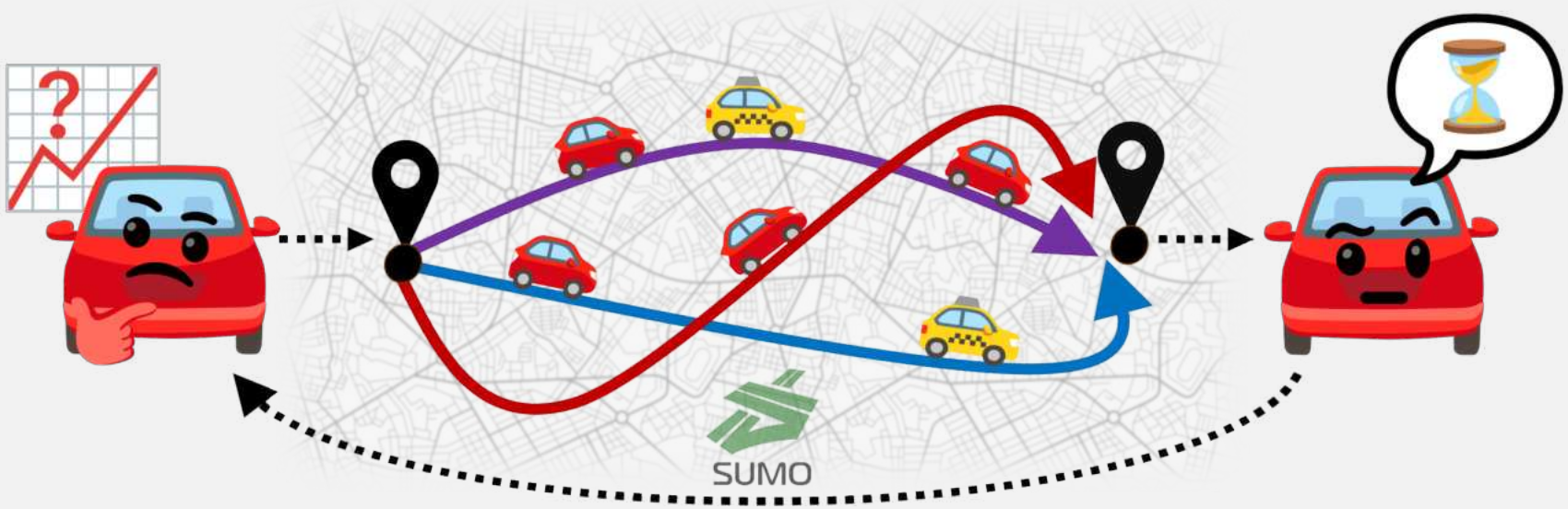
- The traffic system is **shared by humans** and **autonomous vehicles**.
- Humans try to **maximize their individual utilities**, and AVs are controlled by (local or centralized) **automated decision-makers**.
- In such a setting,
  - Are you better off as an AV user?
  - How are humans affected by this coexistence?
  - Is it more favorable for the system's welfare?



Problem

Day-to-day route choice

- A route is a sequence of links that connects a given origin o to a destination d in a network $G = (N, A)$.
- Every day, a traveler i selects a route a_i from their discrete choice set C_{od}, which is the set of routes connecting the traveler's origin and destination.
- In the end, traveler i's realized travel time is the sum of link times on their chosen route, each determined by others' choices and exogenous conditions. ($T_{i(a,z)} = \sum_{a \in A} M_{a,a_i}\, c_a(x_a(\boldsymbol{a}), z_a)$)
- Iteratively, traveler i refines their route-choice policy by updating expected costs from experience to make better decisions next day.



Landscape

Multi-agent reinforcement learning for Markov routing games: A new modeling paradigm for dynamic traffic assignment

Shou, Z., Chen, X., Fu, Y., & Di, X.

Transportation Research Part C: Emerging Technologies (2022)

Can **dynamic routing** among many **selfish drivers** be modeled as a **Markov routing game** and be solved by **MARL**?

Setting

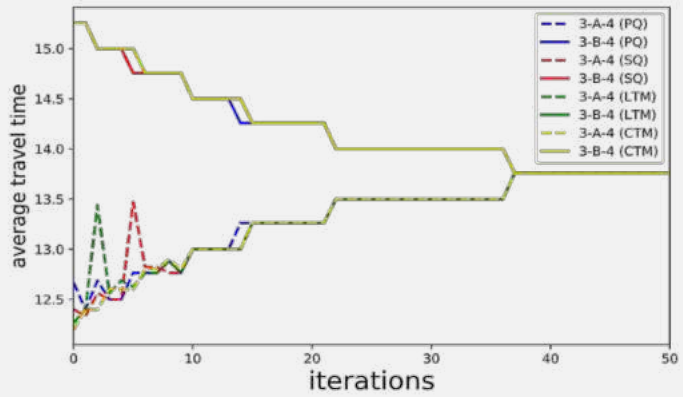
- A variety of traffic networks with varying scales.
- **En-route** path choice.
- Selfish agents competing for link capacities.

Method

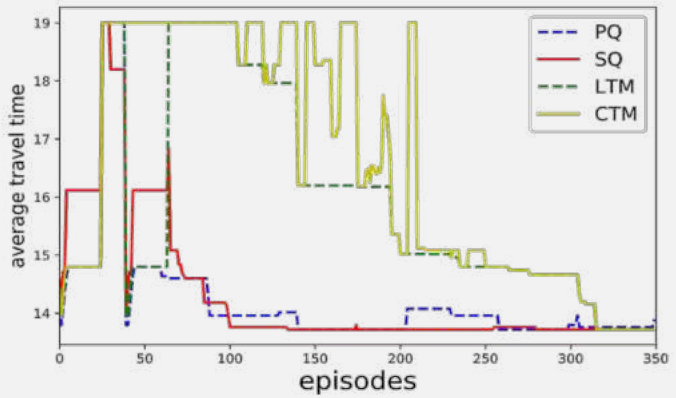
- Model the route assignment as a Markov routing game.
- A **multi-agent mean-field DQL** algorithm for optimal route choice.

Findings

- The learned Markov equilibria are consistent with predictive DUE.
- The approach is computationally efficient on mid-sized and large networks.



(a) The iterative method



(b) MF-MA-DQL algorithm for DUE

Social implications of coexistence of CAVs and human drivers in the context of route choice

Jamróz, G., Akman, A. O., Psarou, A., Varga, Z. G., & Kucharski, R.

Scientific Reports (2025)

In a traffic **bottleneck**, what happens when we deploy **coordinated CAVs** into a **human-only system**?

Setting

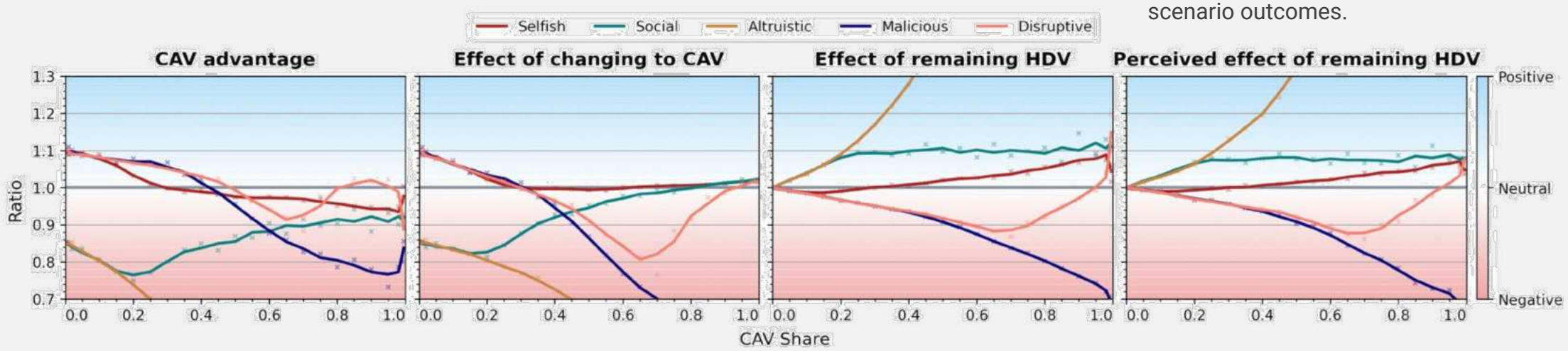
- A traffic network with **two routes connecting an origin to a destination.**
- **Mixed** system.
- Humans are modeled with discrete choice models.

Method

- AVs collectively choose routes according to **their human choice predictions**.
- Rewards are defined according to **behavioral objectives**.

Findings

- CAV choices differ significantly from the choices of the remaining humans.
- Some CAV strategies may result in significant deterioration of driving conditions for all drivers.
- CAV share significantly impacts the scenario outcomes.



Learning how to dynamically route autonomous vehicles on shared roads

Lazar, D. A., Bıyık, E., Sadigh, D., & Pedarsani, R.

Transportation research part C: emerging technologies (2021)

Can deep RL route AVs in **mixed** traffic to guide the system toward the **best traffic equilibrium**, and **reduce congestion**?

Setting

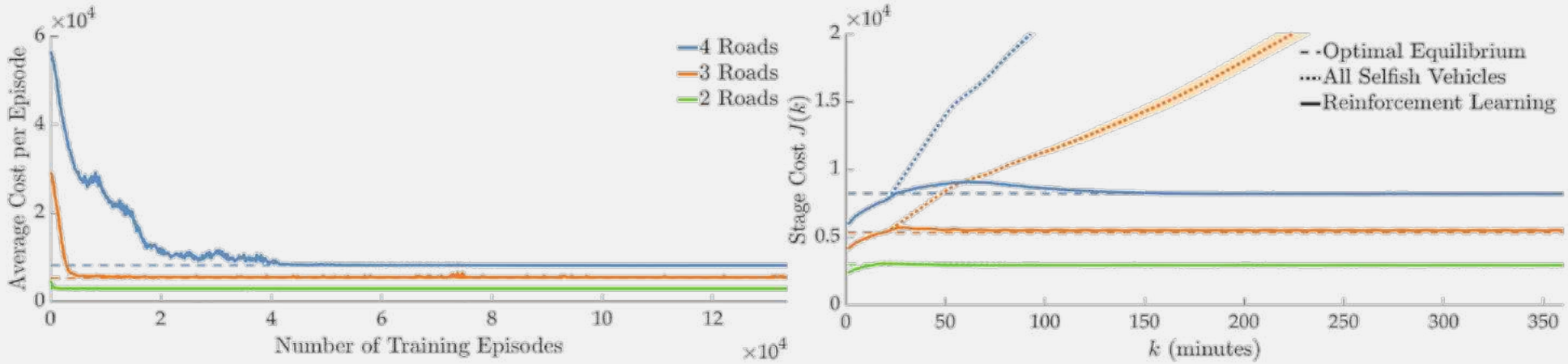
- Simulated networks using a **cell transmission model**.
- Humans keep 2s headway, AVs 1s; 40M training steps.
- Networks include **stochastic demand and accidents**.

Method

- Routing task as a POMDP.
- AV controller trained using **PPO**.
- Reward function: **number of cars in the system** (negated).

Findings

- AV routing can indirectly influence human drivers.
- RL consistently drives the system to near-optimal equilibria.
- RL outperforms static equilibrium-based routing in stochastic settings.



Impact of Collective Behaviors of Autonomous Vehicles on Urban Traffic Dynamics

Akman, A. O., Psarou, A., Varga, Z. G., Jamróz, G., & Kucharski, R.

Seventeenth European Workshop on Reinforcement Learning (2024)

What happens when AVs enter an urban traffic system shared with humans and follow some behavioral objectives?

Setting

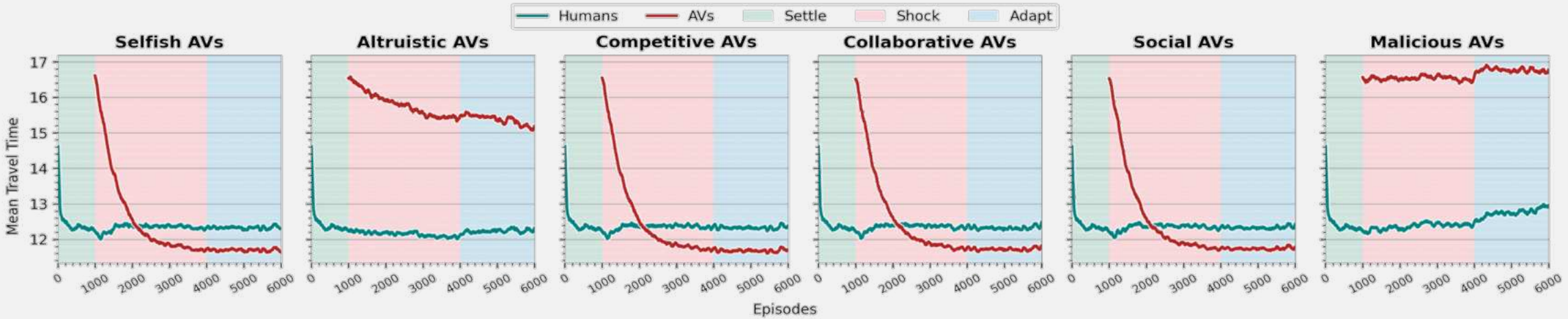
- Real-world traffic network.
- **Mixed** system.
- Humans are modeled with **discrete choice models**.

Method

- AVs trained using **IQL**.
- Rewards are defined according to **behavioral objectives**.
- Training mimics a deployment scenario, divided into **phases**.

Findings

- AVs achieve better travel times when they aim for it.
- AV deployment causes human preference shifts.
- In most cases, humans are disadvantaged.



Impact of Collective Behaviors of Autonomous Vehicles on Urban Traffic Dynamics

Akman, A. O., Psarou, A., Varga, Z. G., Jamróz, G., & Kucharski, R.

Seventeenth European Workshop on Reinforcement Learning (2024)

Setting

- Real-world traffic network.
- **Mixed** system.
- Humans are modeled with **discrete choice models**.

Method

- AVs trained using **IQL**.
- Rewards are defined according to **behavioral objectives**.
- Training mimics a deployment scenario, divided into **phases**.

JAGIELLONIAN UNIVERSITY IN KRAKÓW

COeXISTENCE

JAGIELLONIAN UNIVERSITY IN KRAKÓW

erc

Impact of RL-enabled autonomous vehicle route choice behaviors on urban traffic dynamics

Ahmet Onur Akman*, Anastasia Psarou, Zoltán György Varga, Grzegorz Jamróz, Rafał Kucharski

Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland

RL-enabled autonomous vehicles can optimize behavioral objectives by learning to choose better routes!

And this may have a noticeable impact on human drivers and traffic efficiency.

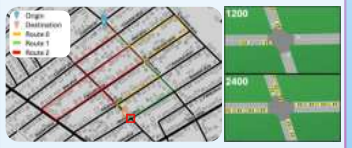
Introduction

- We investigate the coexistence of **human drivers** and **RL-enabled autonomous vehicles** (AVs) in a **day-to-day route choice** scenario.
- Every day, drivers **pick one of the routes** connecting their origin-destination (OD) pairs. Day by day, they refine their route preferences according to **experienced travel times**.

- Each AV agent uses individual DQNs and learns from individual experiences. Human agents are modeled with a state-of-the-art behavioral model.
- We explore the scenarios in which AVs adopt different **behavioral objectives**.
- Our experiments are conducted using our MARL framework: **PARCOUR**.

Problem

- Each episode simulates the commute of **1000 drivers** at a **rush hour** in traffic.
- There are two origins, two destinations, and **three predefined routes** for each of the **four ODs**.
- An episode consists of a single-step decision of each agent, which is their route choice for the day.
- **Humans seek to minimize individual travel times.** AVs aim to optimize their returns.



Left: Csömör traffic network used in our experiments, with routes connecting OD (0,0).
Right: The congestion at the highlighted intersection on indicated timesteps within an episode.

AV Behaviors

- We define six behaviors for the AVs with different reward formulations:

- The reward function is a linear combination of the mean travel times of different subsets of agents, **weighted by the behavioral coefficients**.
- In each experiment, all AVs **uniformly** adopt a selected behavior.

Experimental Setting

Phase I: Settle
There are **only human drivers** in the traffic. They learn about the environment and tune their cost expectations to make better route choices.

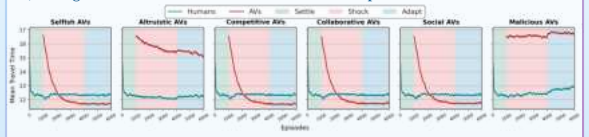
Phase II: Shock
AVs replace 1/3 of the vehicles. Now AVs optimize their policies, while the remaining **humans stick to their learned preferences.**

Phase III: Adapt
Humans are now reacting to the changes, by updating their preferences. This creates an environmental shift, and **AVs adjust their policies.**

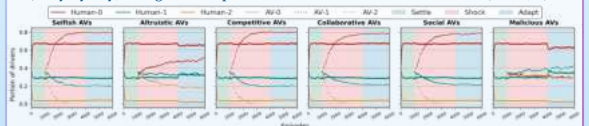
Days (Episodes)

Results

1) Changes in human and AV travel times over the episodes for each AV behavior.



2) Day by day, changes in the preferences of human drivers and AVs. (OD 0, 0)



3) Consequences of AV deployment in each scenario. We compare the **mean travel times** of **humans at the end of Phase 1** and **the indicated group at the end of Phase 3**. A positive effect indicates reduced travel time.

Behavior \ Effect on	AV Travel Time	Human Travel Time	Traffic Efficiency
Altruistic	-23.1%	+0.3%	-7.1%
Collaborative	+4.3%	-0.7%	+0.9%
Competitive	+4.6%	-0.8%	+0.9%
Malicious	-36.3%	-5.4%	-15.1%
Selfish	+4.6%	-0.7%	+1.0%
Social	+4.2%	-0.7%	+0.9%

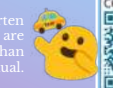
Takeaways

AV users enjoyed shorter commutes with each self-travel-time optimizing AV behavior.

In most cases, the AV deployment improved the traffic efficiency.

Malicious and altruistic behaviors caused great delays for AV users.

When AVs aim to shorten their travels, drivers are better off switching than remaining manual.



This research is financed by the European Union within the Horizon Europe Framework Programme. ERC Starting Grant COEXISTENCE no. 101075838

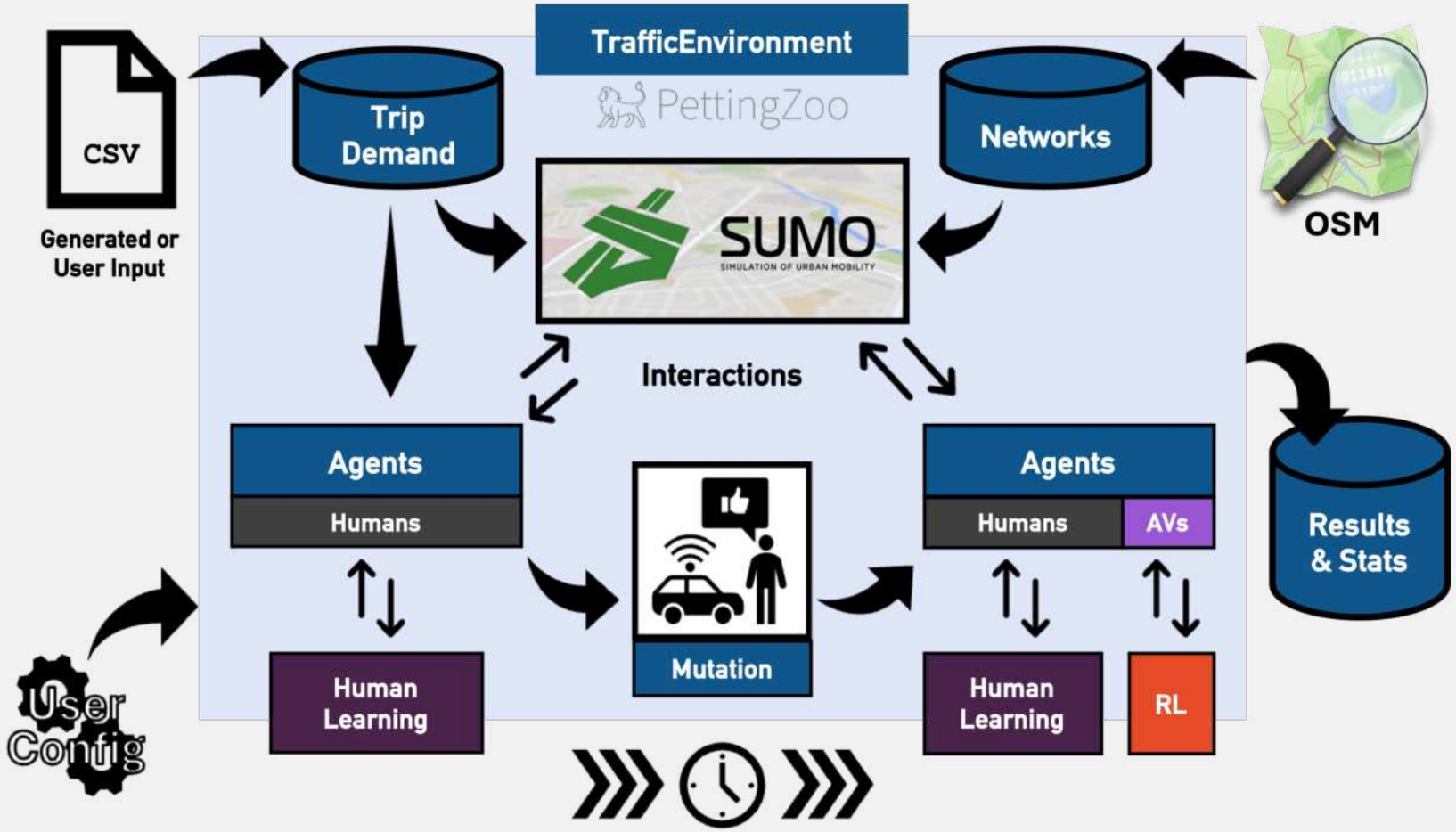
EXPLORE COeXISTENCE

Contributions

RouteRL

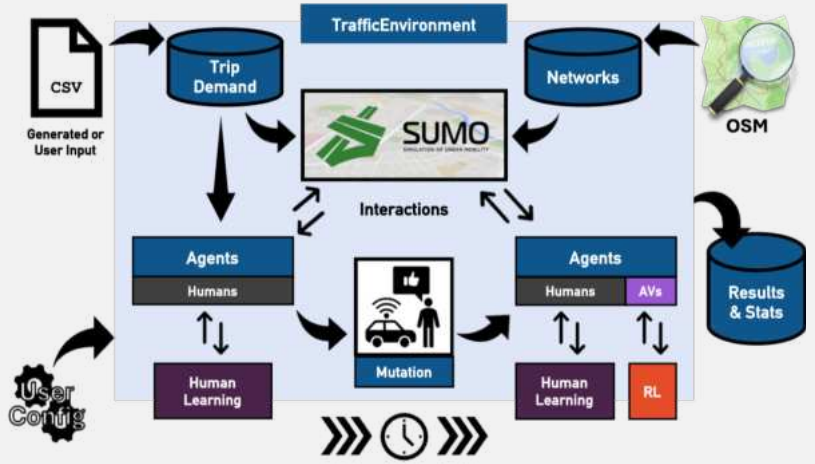
- RouteRL bridges RL-based routing with a microscopic traffic simulation.
- It facilitates:
  - Analysis of the potential impact of AVs in future cities,
  - Testing novel (MA)RL solutions in multi-agent route choice,
  - Empirical validation for projections on future traffic dynamics with AVs.

RouteRL


TrafficEnvironment
PettingZoo
CSV
Generated or User Input
Trip Demand
Networks
OSM
SUMO
SIMULATION OF URBAN MOBILITY
Interactions
Agents
Humans
Agents
Humans
AVs
Results & Stats
Mutation
User Config
Human Learning
Human Learning
RL

RouteRL

- RouteRL* enables experimenting with:
  - Custom or SOTA (MA)RL algorithms,
  - Human learning and decision-making models,
  - Custom scenarios,
  - Traffic network topologies.



Paper

Code

* Akman, A. O., Psarou, A., Gorczyca, Ł., Varga, Z. G., Jamróz, G., & Kucharski, R. (2025). RouteRL: Multi-agent reinforcement learning framework for urban route choice with autonomous vehicles. *SoftwareX*, *31*, 102279.

URB

Urban Routing Benchmark

- URB is a comprehensive **benchmarking framework** powered by RouteRL, aiming to **standardize the assessment** for AV routing solutions.
- It unifies evaluation across **real-world inspired tasks**.
- It comes with a catalog of implementations, baselines, domain-specific performance indicators, and a reusable configuration scheme.

URB

Dataset

Task variety

A diverse set of scenarios to test solution robustness under different complexities.

29 real-world networks

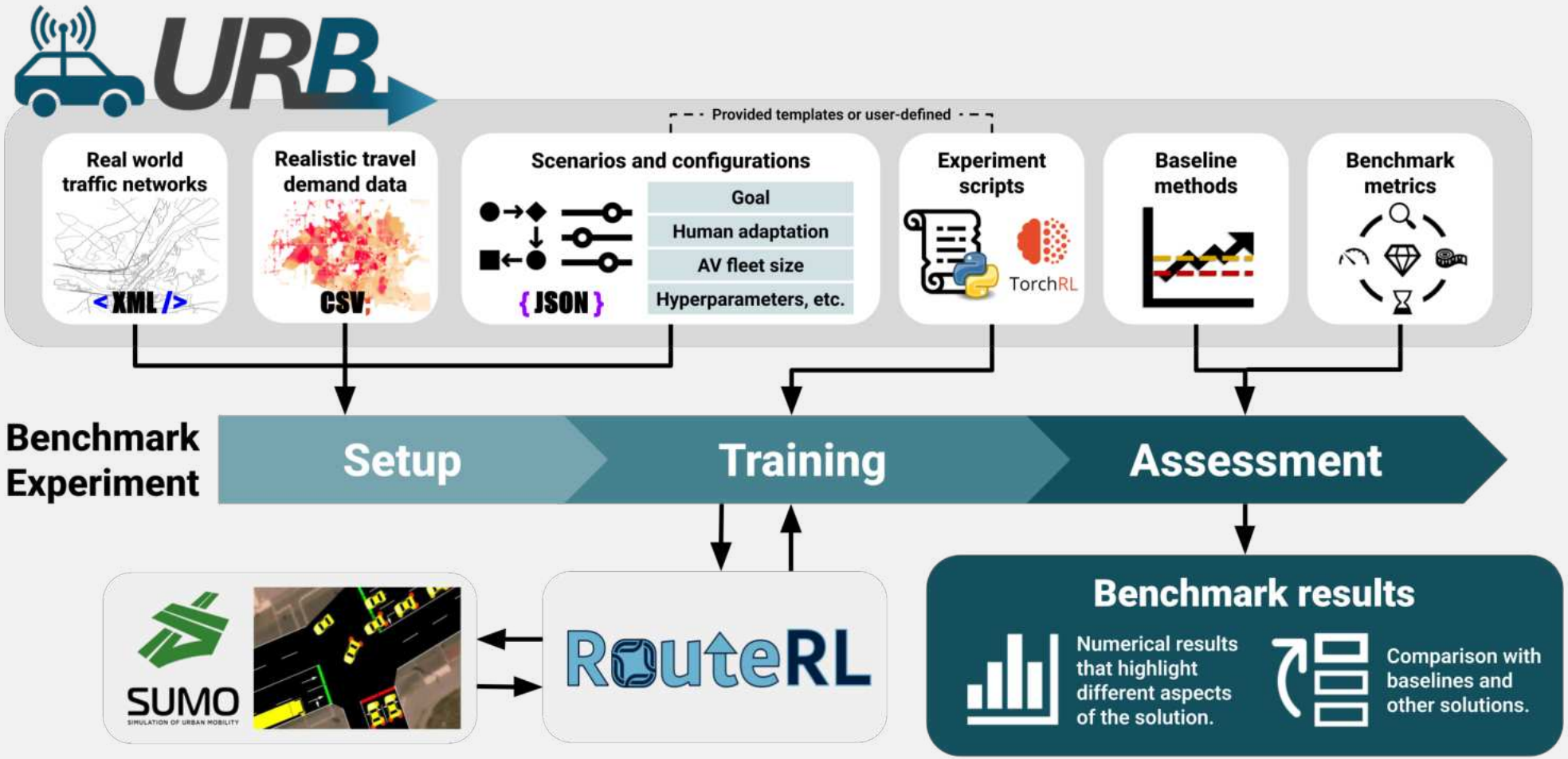
28 towns from the Île-de-France region, and Ingolstadt from Germany.

Realistic demand data

Each network is bundled with realistic demand data for assessment in realistic scenarios.

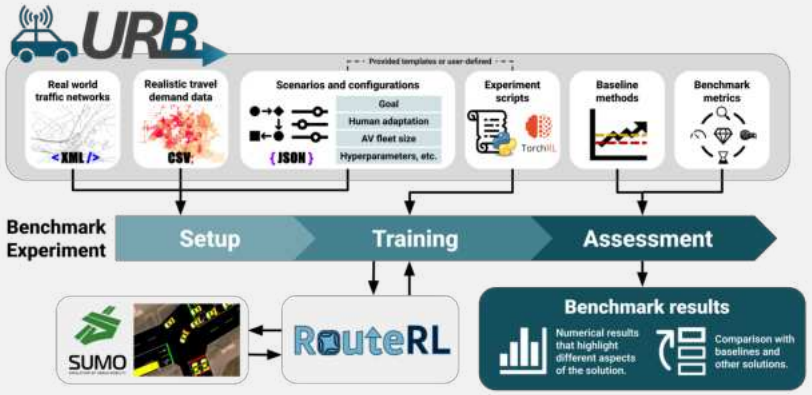
Open access

The dataset and its generation code are documented and made available publicly.


URB
Provided templates or user-defined
Real world traffic networks
<XML />
Realistic travel demand data
CSV;
Scenarios and configurations
{JSON}
Goal
Human adaptation
AV fleet size
Hyperparameters, etc.
Experiment scripts
TorchRL
Baseline methods
Benchmark metrics
Benchmark Experiment
Setup
Training
Assessment
SUMO
SIMULATION OF URBAN MOBILITY
RouteRL
Benchmark results
Numerical results that highlight different aspects of the solution.
Comparison with baselines and other solutions.

URB

- URB* is our first step towards a **standardized evaluation** in urban AV routing for efficient, reliable, and transparent solutions.
- We are actively working on expanding our ***leaderboard*** to increase **methodological diversity**.



Paper

Code

Dataset

* Akman, A. O., Psarou, A., Hoffmann, M., Gorczyca, Ł., Kowalski, Ł., Gora, P., ... & Kucharski, R. (2025). URB--Urban Routing Benchmark for RL-equipped Connected Autonomous Vehicles. *arXiv preprint arXiv:2505.17734*.

Significance

Significance

People and Policies

- **Commute & Quality of Life:** The way AVs choose routes directly influences **travel times, congestion, and stress in daily mobility**.
- **Efficiency & Sustainability:** Smarter route choices can reduce **emissions, fuel use, and wasted time** on the roads.
- **Safety & Fairness:** Outcomes determine whether AVs can be used for the system's welfare and ensure that **humans are not disadvantaged**.
- **Policy & Regulation:** **Evidence is needed to set rules** that prevent harmful AV behaviors and promote socially beneficial ones.

Significance

RL Researchers

- **Complex Multi-Agent Setting:** The setting offers a **realistic testbed** for MARL, where a **large group of agents** (in the scale of hundreds to thousands) interact, adapt, collaborate and/or compete over time.
- **Human-AI Coexistence:** Bridges behavioral models of humans with RL agents, highlighting challenges of **mixed autonomy**.
- **Non-Stationary Environments:** Provides a natural domain where policies must **adapt to evolving human and system behaviors**.
- **Societal Relevance:** Connects algorithmic advances to tangible societal outcomes, **strengthening the impact and visibility of RL research**.

sim2real

Lifelong Learning

Meta-RL & Transfer Learning

Can our solutions be system-agnostic or adaptive to topologies and patterns without complete retraining?

Equilibrium Learning

Inverse RL

Can we infer drivers' latent utilities and adaptation rules from public datasets to anticipate and respond to human learning dynamics?

Ad Hoc Teamwork

Can routing policies remain effective as fleet size or partners change without prior coordination?

Risk-Sensitive RL

Can exploration and deployment satisfy risk budgets under rare but critical events?

Hierarchical RL

Can high-level policies be effectively paired with low-level decisions to achieve scalability in large systems?

Incentive Shaping

Can we achieve social optimality with learned tolls, information signals, or priority rules?

Cooperative Exploration

How can large AV fleets explore useful states in large spaces in real time at minimal cost?

Communication Under Budget

Thank you!

/aonurakman

rafalkucharskilab.pl/COeXISTENCE/

Research financed by the European Union within the Horizon Europe Framework Programme (ERC Starting Grant COeXISTENCE no. 101075838).