08/16/2025

Autonomous Driving

Introduction, Technologies, and the Planning Problem

Aleksandr Petiushko, https://petiushko.info
Head of AI Research @ Elea

RnD @ IT Picnic

Contributors

Ashwin Balakrishna

Jonathan Booher

Ishan Gupta

Vladislav Isenbaev

Wei Liu

Aleksandr Petiushko

Khashayar Rohanimanesh

Junhong Xu

*Mentioned in the alphabetical order

AD and SDV

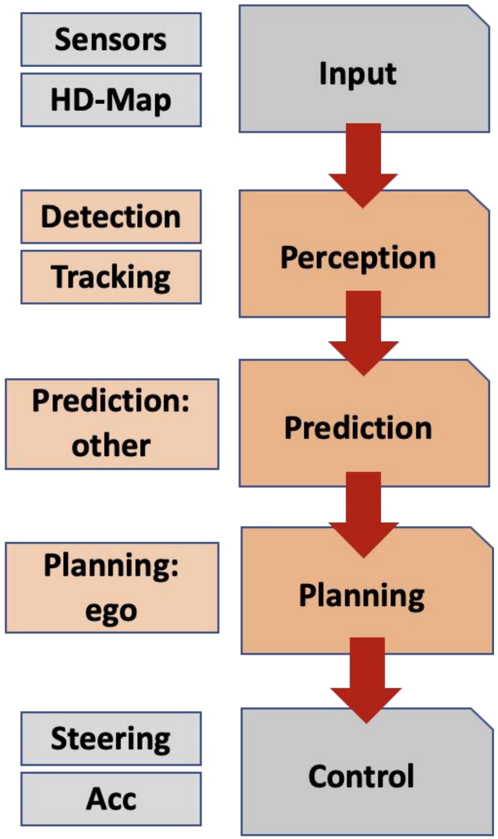
- **AD** = Autonomous Driving: the *task*
- **SDV** = Self-Driving Vehicle: the *car*
- *AD* is one of the most complex and difficult tasks, both theoretically and practically

Image source

Safety of SDV and other agents on the road is crucial

AD: Classical ML Stack of Technologies

- The main **software** parts are the so-called **P**^{3}:
  - Perception, Prediction and Planning
- **Hardware** parts:
  - Input: Sensors
  - Output: Control (steering, acceleration)
- High-Definition Map as the helper
  - **HD-Map** contains info about the road



SDV: Sensors

- Various **sensors** are used:
  - LIDAR
  - Radar
  - Ultrasonic
  - Cameras (*x N*)

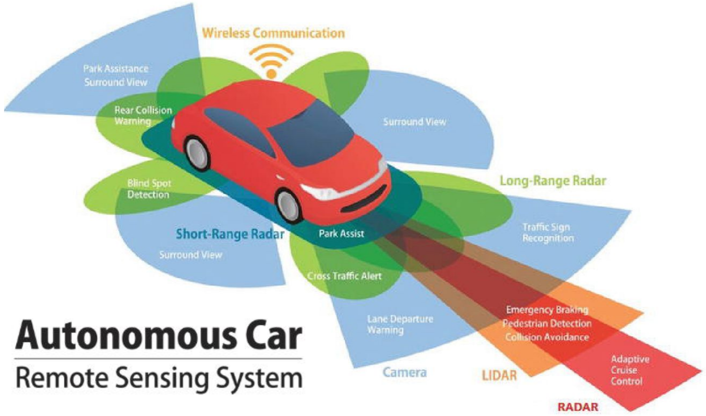
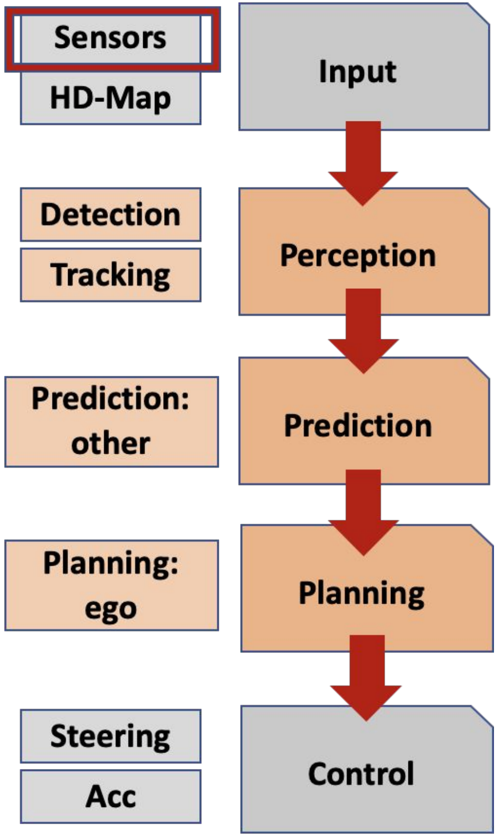


Image [source]

- **Problems**:
  - Expensive
  - Hard to synchronize



AD: HD-Map

- Helpful for prediction and planning
  - Contains information about a **road**:
    - Lanes, crosswalks, traffic lights, etc.
- **Problems**:
  - Every company has its own format
  - Significant overhead

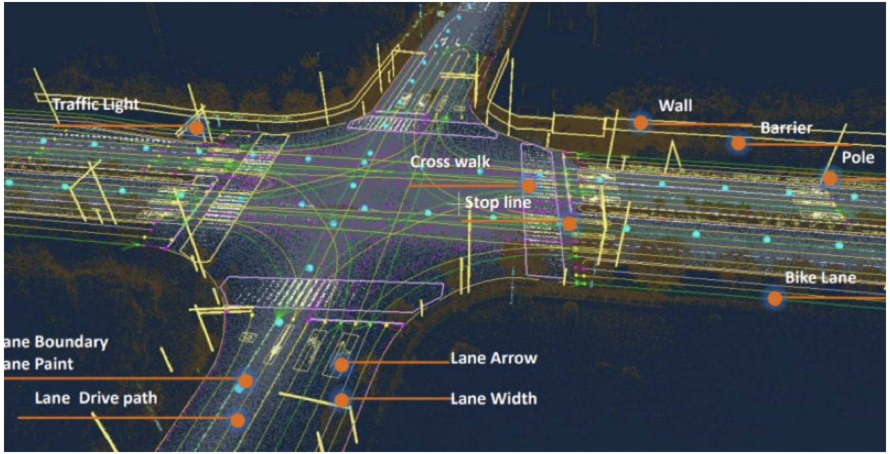
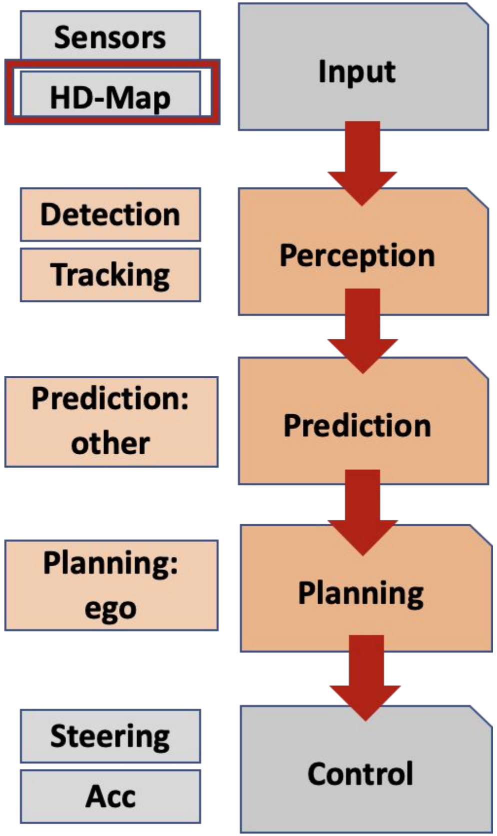


Image [source]()



AD: Detection

- The *first* step of the Perception part:
  - **Detection** (segmentation, depth-estimation, etc.) of the objects around
- **Problems**:
  - Long tail (small and unusual objects) and anomalies

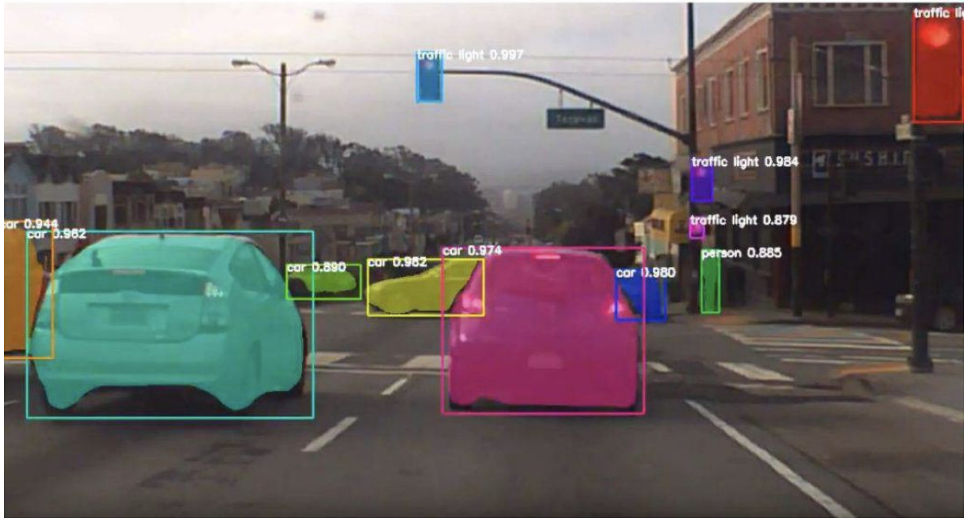
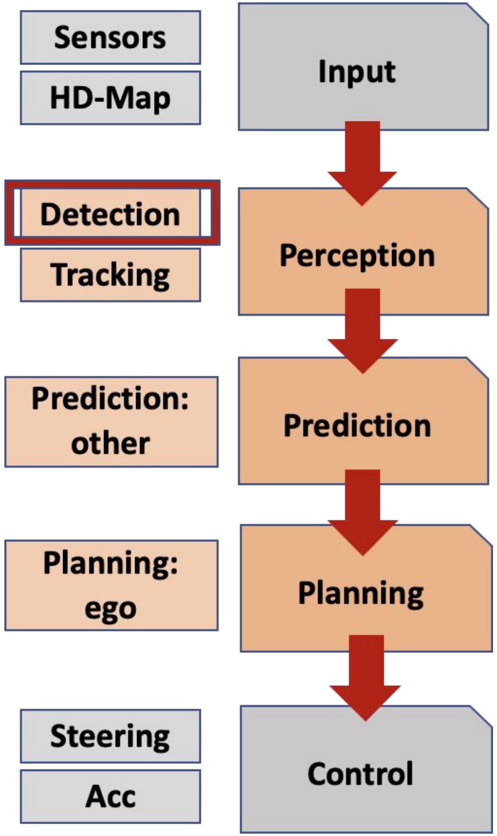


Image source



AD: Tracking

- The *second* step of the Perception part:
    - **Tracking** of the detected objects and estimation of their coordinates for the Prediction part
- **Problems**:
    - Track association of flickering objects

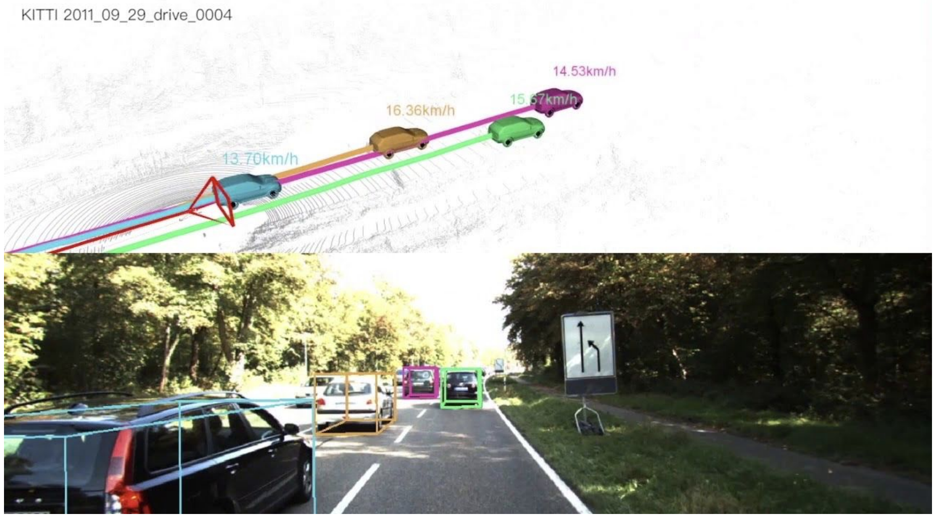
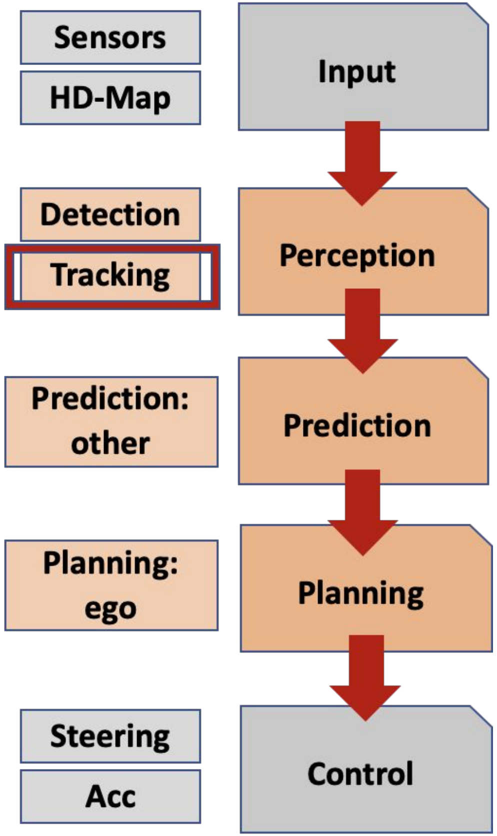


Image [source]



AD: Prediction

- Future trajectories **prediction** of all surrounding objects based on the *tracking history* and *HD-Map*
    - Usually, 1-10 second
- **Problems**:
    - Multi-modality for recall

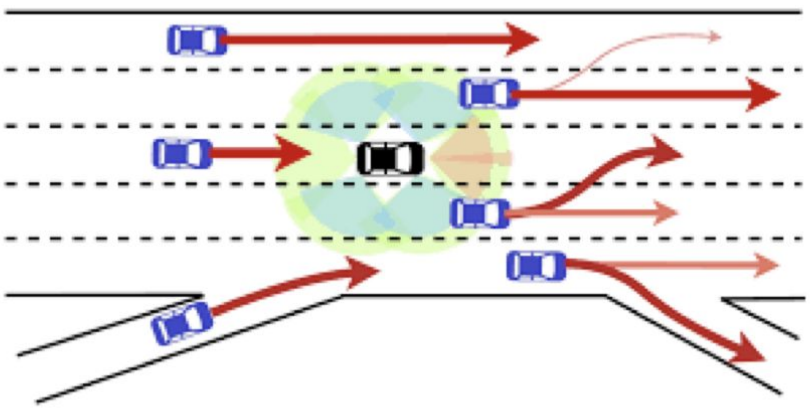
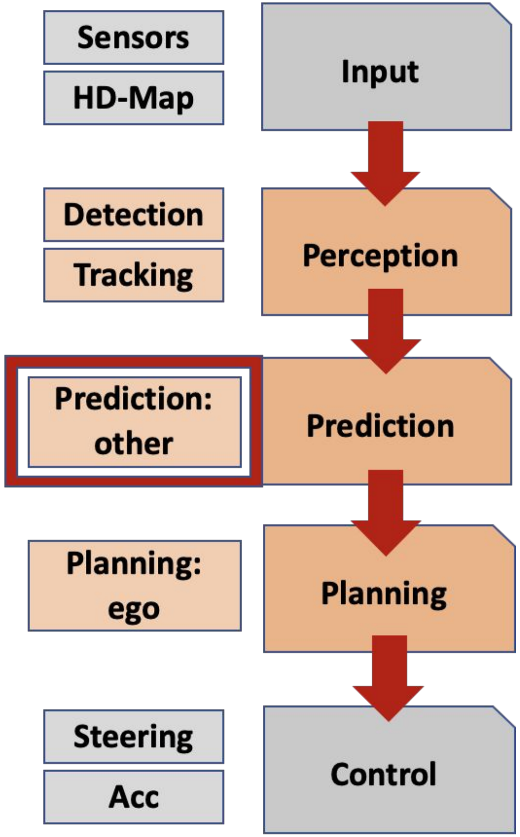


Image source



AD: Planning

- **Planning** of SDV future actions based on the *predictions* and *HD-Map*
- **Problems**:
    - Consistent joint prediction and planning

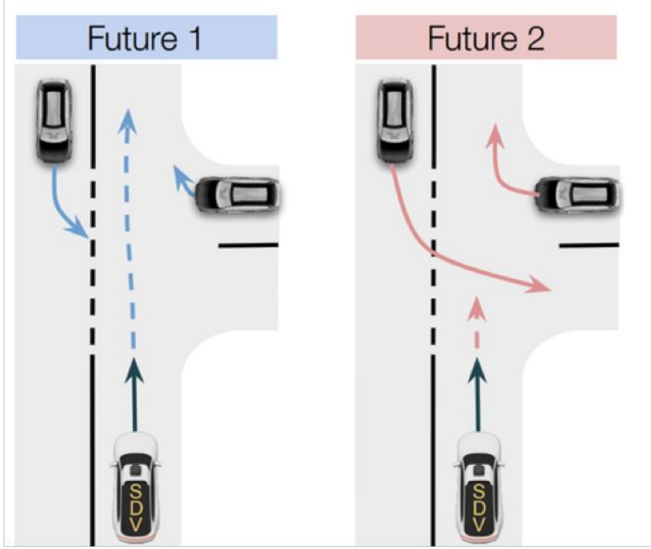
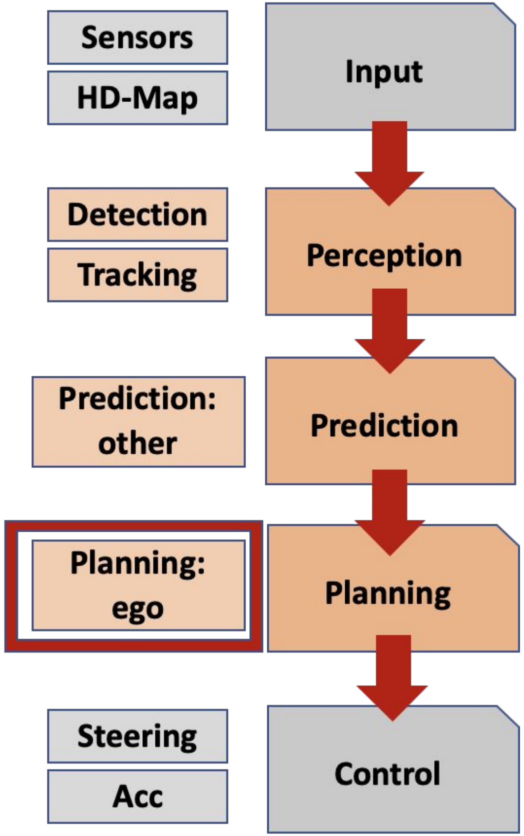


Image [source]



SDV: Control

- Realization and **control** of SDV actions based on *motion plan*
  - Steering control, acceleration control, etc.

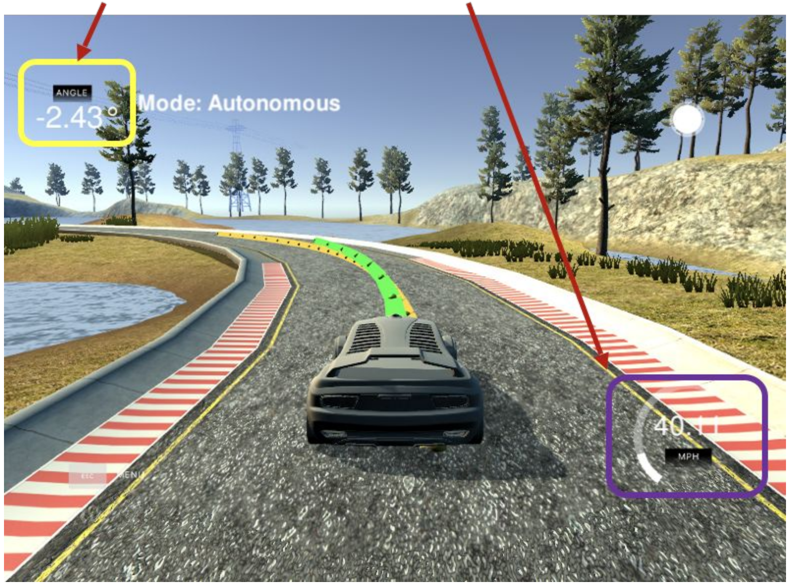
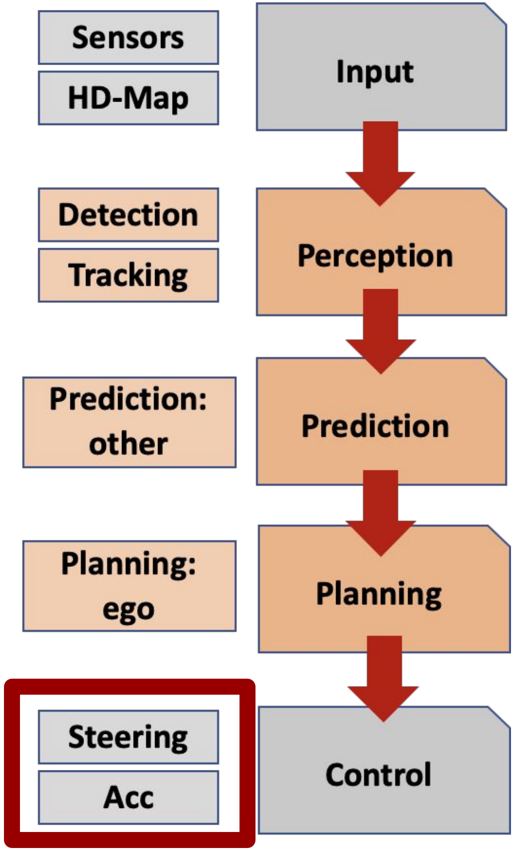


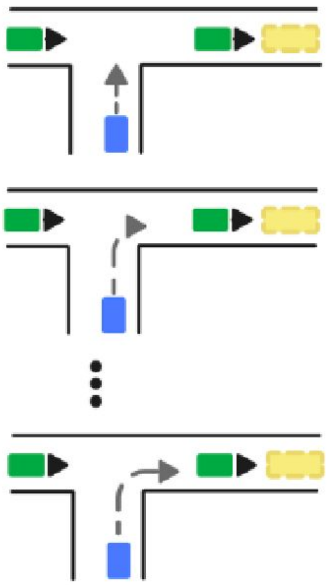
Image source

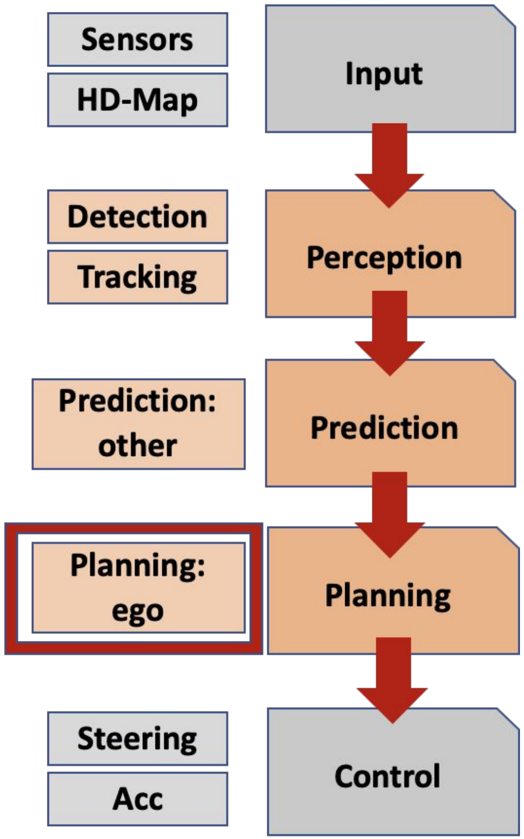
- **Problems**:
  - Dynamic and kinematic limitations



How to choose the right plan?

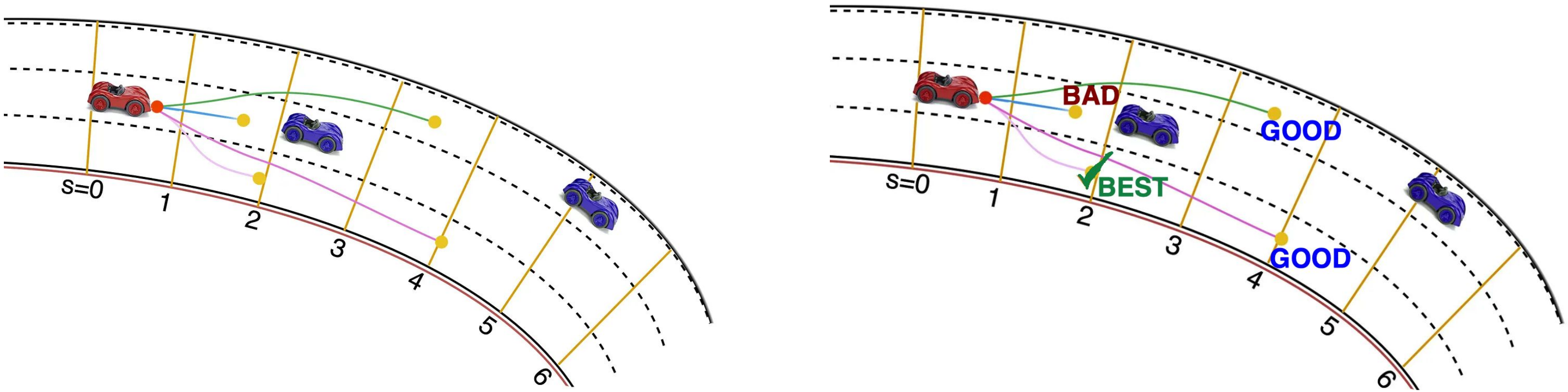
- Need a scorer!





Plan Generation vs Plan *Selection*

Generation **Selection**



Plan Generation vs Plan Selection (Image source)

Let's **combine** two worlds!

Imitation Learning

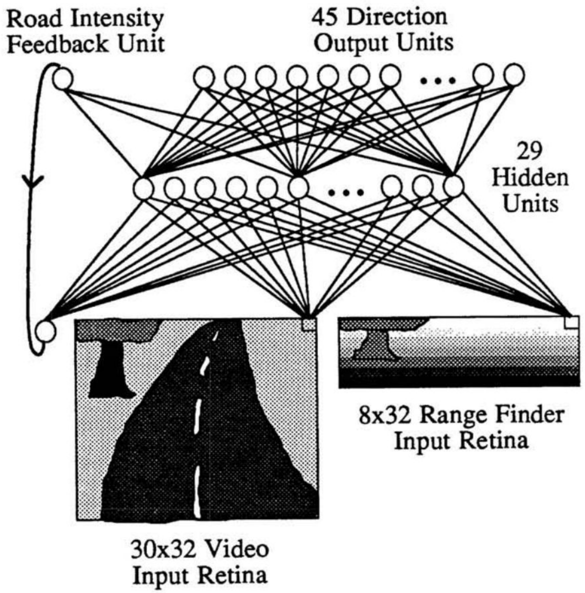


Figure 1: ALVINN Architecture

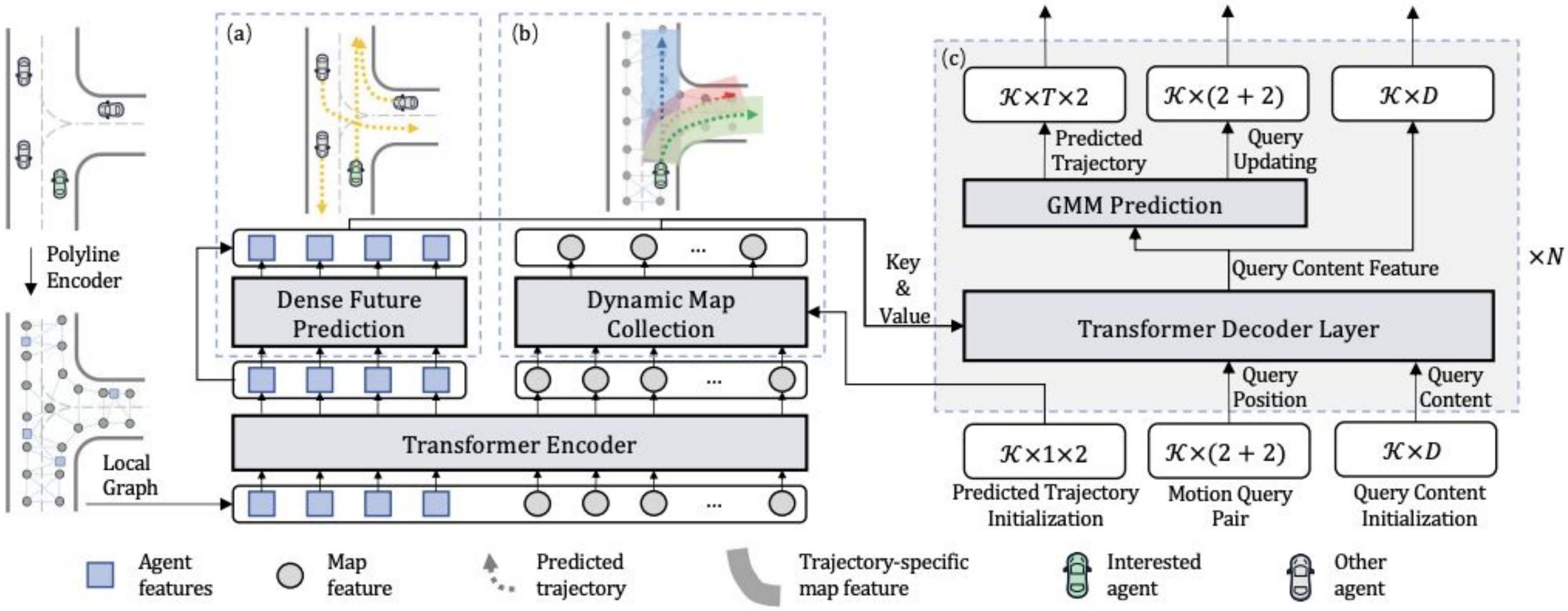
"NN can accurately drive the Ego Vehicle at a speed of 1/2 mps along a 400 m path through a wooded area under sunny fall conditions."

– Behavior Cloning from 1988 (!)

Pomerleau, Dean A. "Alvinn: An autonomous land vehicle in a neural network." 1988.

Imitation Learning

SotA Prediction model:
Motion TRansformer (MTR and MTR++)



Shi, Shaoshuai, et al. "Motion transformer with global intention localization and local movement refinement." 2022.
Shi, Shaoshuai, et al. "MTR++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying." 2023.

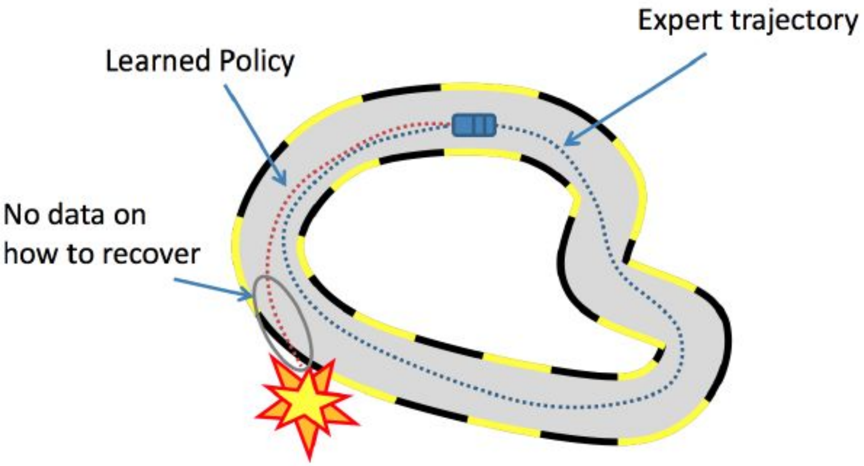
Imitation Learning

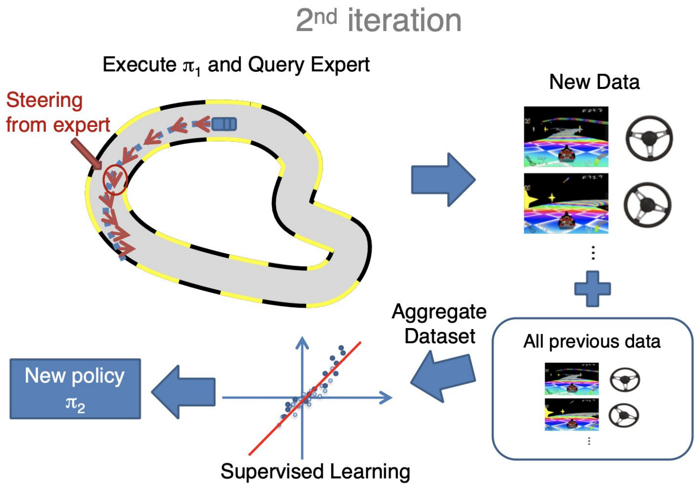
Pros:

- ➔ Simple constructive algorithm scaling with data

Cons:

- ➔ Hard to stay “in distribution” (error quickly accumulates)
- ➔ Can be mitigated by Dataset Aggregation (DAgger) approach

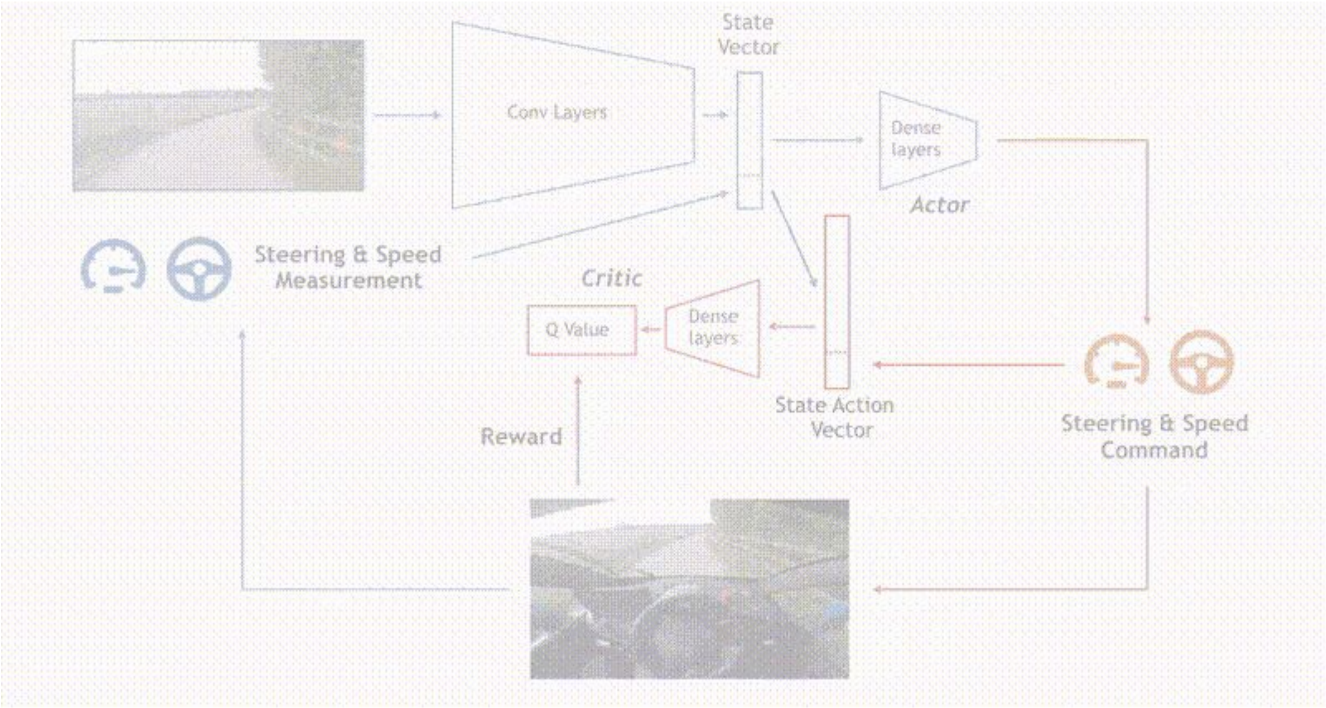




Ross, Stéphane, Geoffrey Gordon, and Drew Bagnell. "A reduction of imitation learning and structured prediction to no-regret online learning." 2011.

Reinforcement Learning

Online, off-policy RL (DDPG) from 2018



Kendall, Alex, et al. "Learning to drive in a day." 2018.

Reinforcement Learning

Pros:

➔ Adaptable to unseen scenarios
➔ Reasoning beyond imitation (hypothetical roll-outs)

Cons:

➔ Hard to define rewards (human-like behavior)
➔ Need reliable infrastructure for reliable estimation at scale

IL+RL

Status Quo:

- ➔ Very good imitation-based models (for Prediction, Planning)
- ➔ Models can be of different nature (ML-based, heuristic-based, simple geometric roll-outs, LLM-based for high-level reasoning, etc)
- ➔ RL policies need to deal with either discretization of the action space or with approximations of the policy gradients

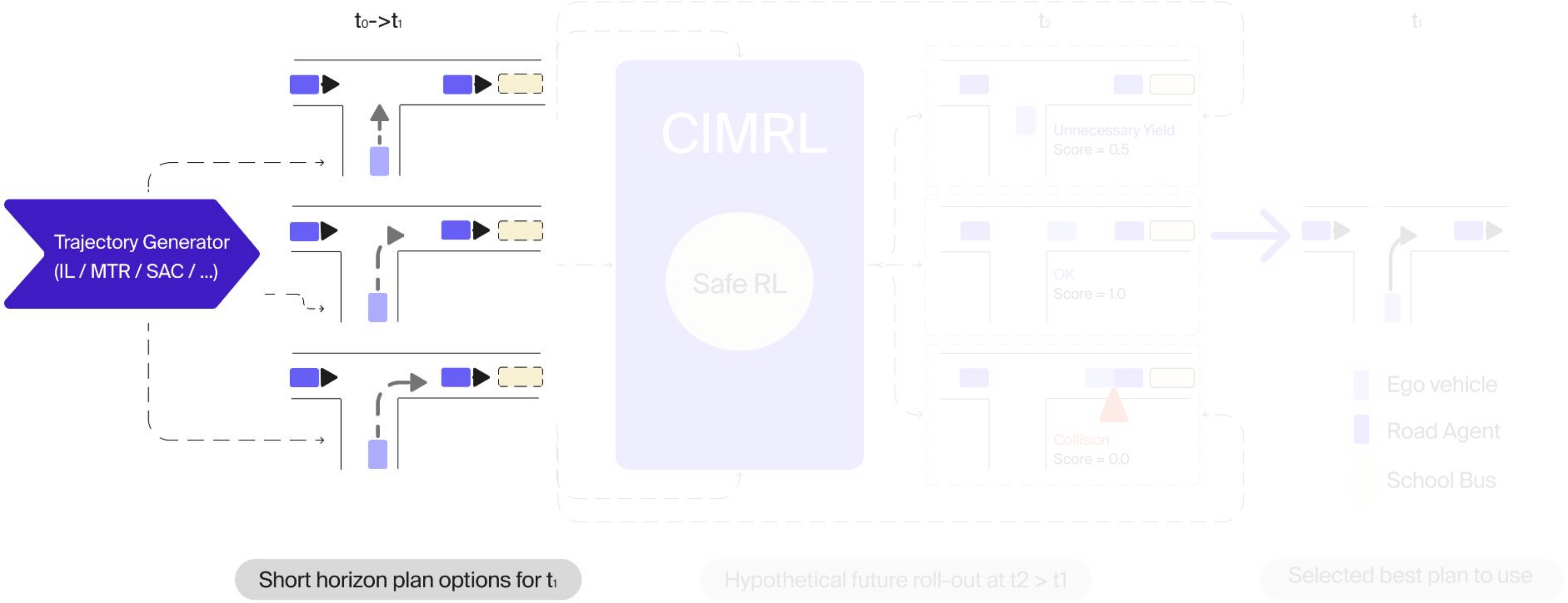
What if:

- ➔ We will re-use the imitation-based existing models, but
- ➔ Use RL algorithm to select from multiple IL generators

Plus:

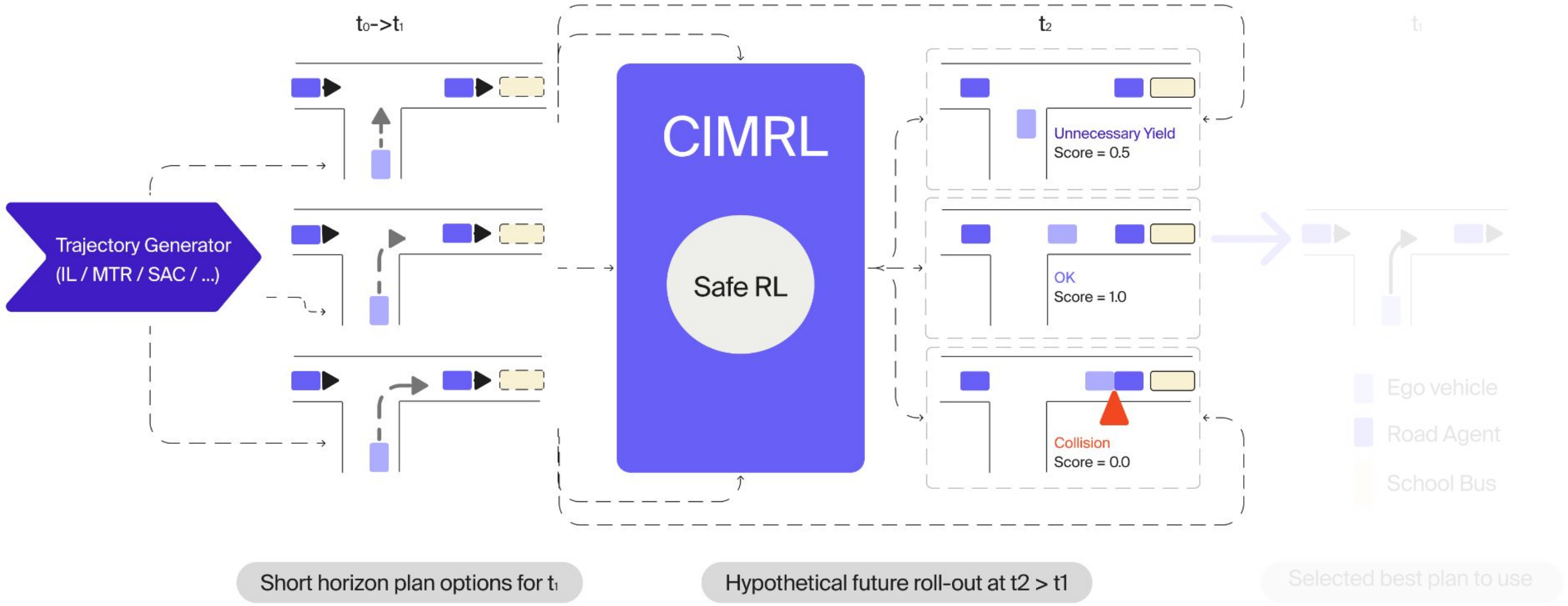
- ➔ We can concentrate on safety by doing hypothetical future roll-outs and remove / downvote dangerous plans, and provide behavior realism from IL

CIMRL: Combining IMitation and Reinforcement Learning



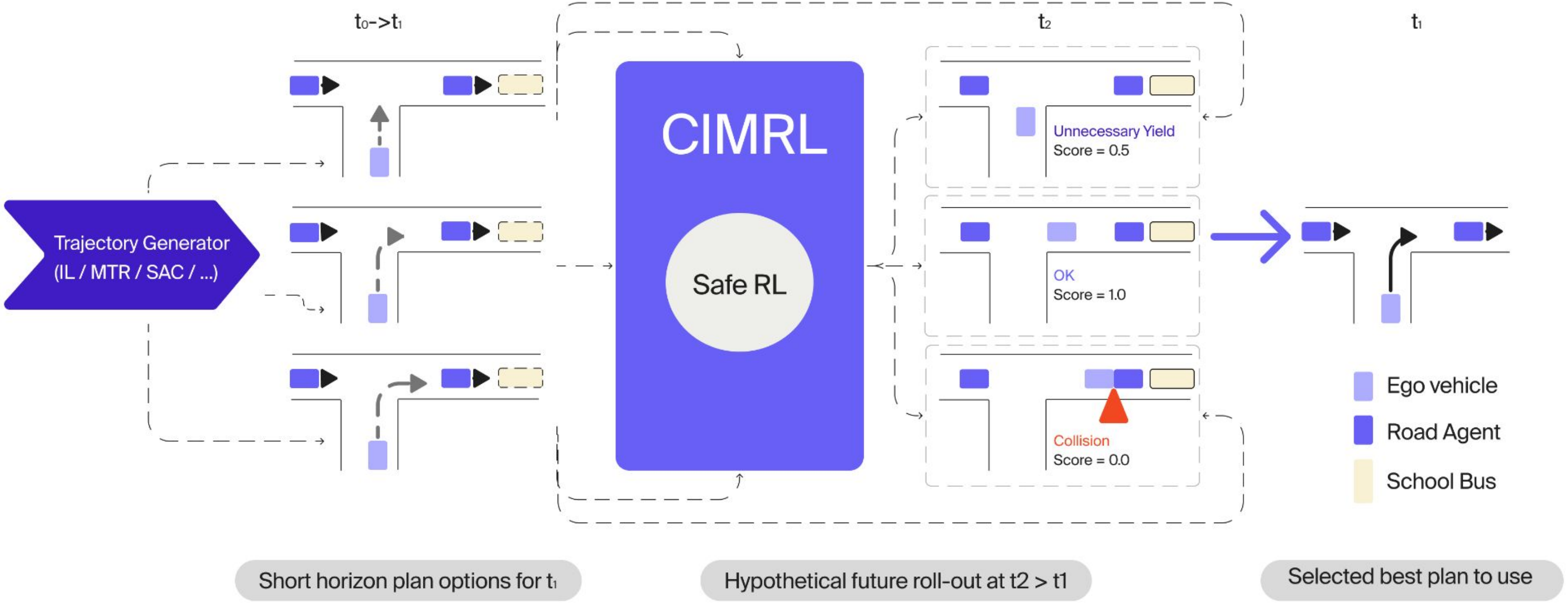
Booher, Jonathan, et al. "CIMRL: Combining IMitation and Reinforcement Learning for Safe Autonomous Driving." 2024.
https://arxiv.org/abs/2406.08878

CIMRL: Combining IMitation and Reinforcement Learning



Booher, Jonathan, et al. "CIMRL: Combining IMitation and Reinforcement Learning for Safe Autonomous Driving." 2024.
https://arxiv.org/abs/2406.08878

CIMRL: Combining IMitation and Reinforcement Learning

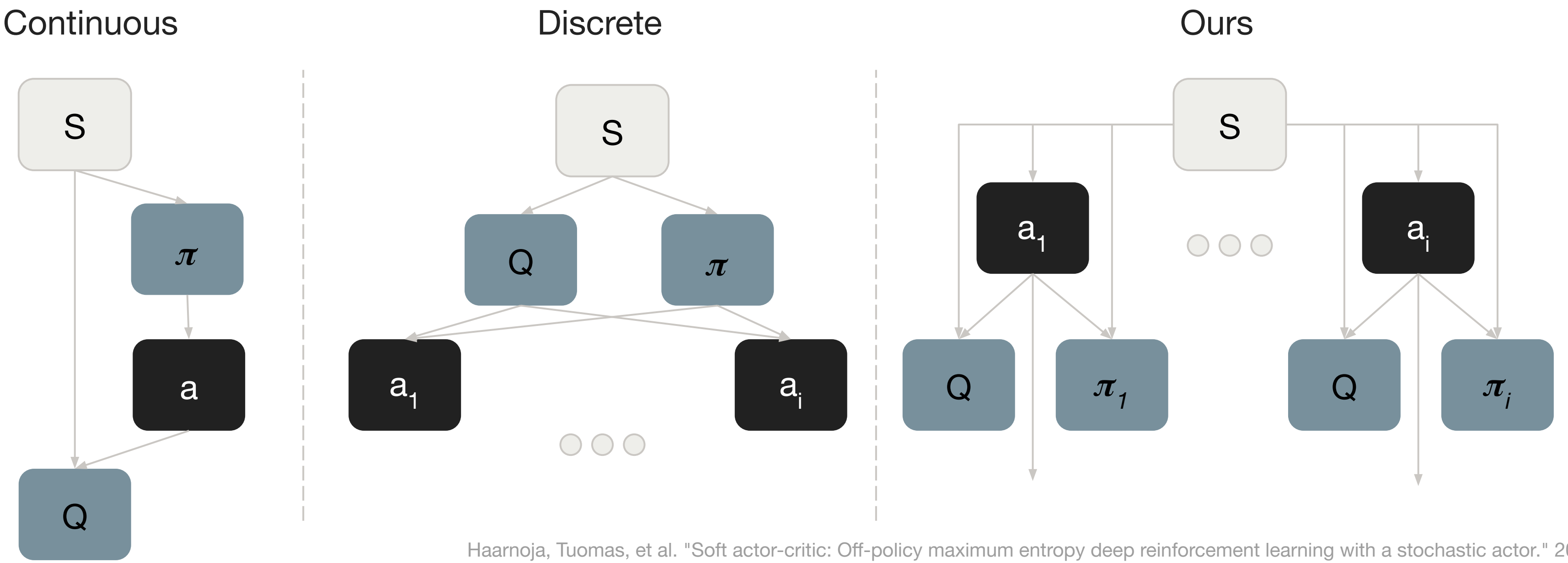


Booher, Jonathan, et al. "CIMRL: Combining IMitation and Reinforcement Learning for Safe Autonomous Driving." 2024.
https://arxiv.org/abs/2406.08878

CIMRL: Scoring

One more (:wink:) combination of:

- ➔ **Continuous** Action Space: able to provide the scoring for literally any planned trajectory
- ➔ **Discrete** Action Space: able to provide the correct probability distribution on top of any finite set of traject



Haarnoja, Tuomas, et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." 2018.
Christodoulou, Petros. "Soft actor-critic for discrete action settings." 2019.

CIMRL: Advantages

Scalability

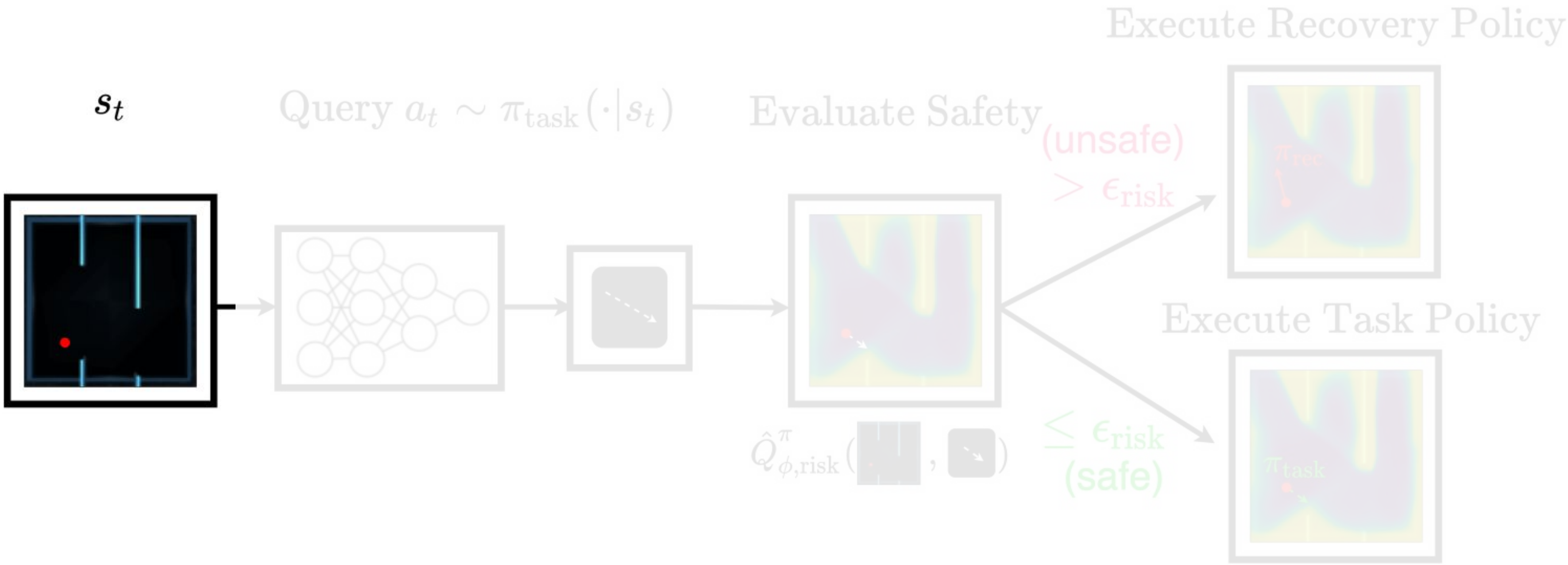
- ➔ Benefits from a lot of data which is directly improving IL-based methods

Flexibility

- ➔ Can be used as a framework for incorporating literally any Prediction or Planning model
- ➔ We can also incorporate the scores from those models as well!

Anatomy of the CIMRL Model: Recovery RL

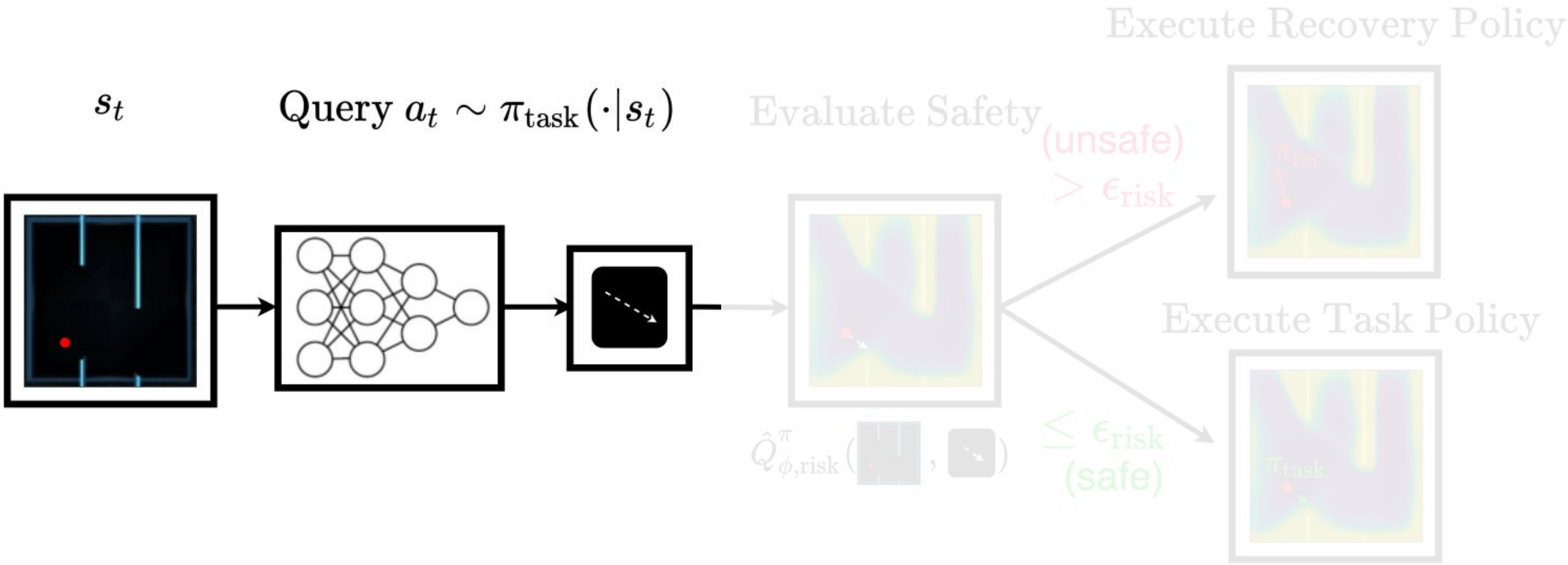
During Inference!



Thananjeyan, Brijen, et al. "Recovery RL: Safe reinforcement learning with learned recovery zones", 2021.

Anatomy of the CIMRL Model: Recovery RL

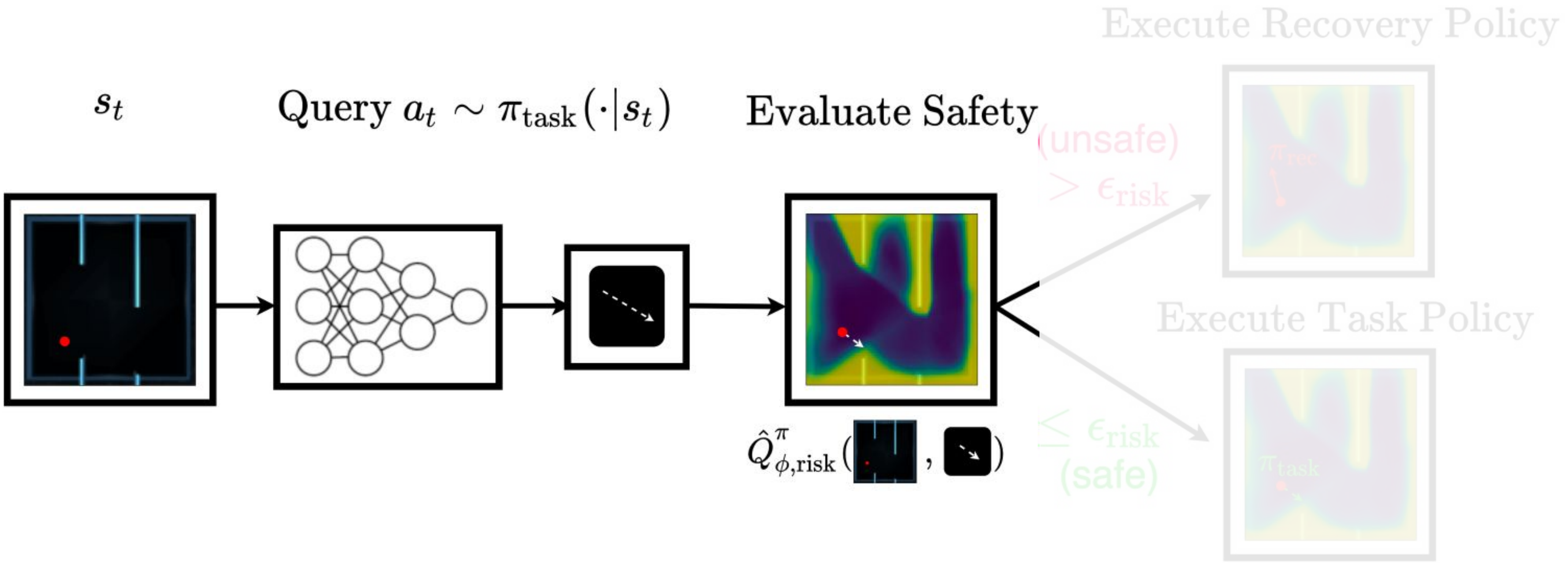
During Inference!



Thananjeyan, Brijen, et al. "Recovery RL: Safe reinforcement learning with learned recovery zones", 2021.

Anatomy of the CIMRL Model: Recovery RL

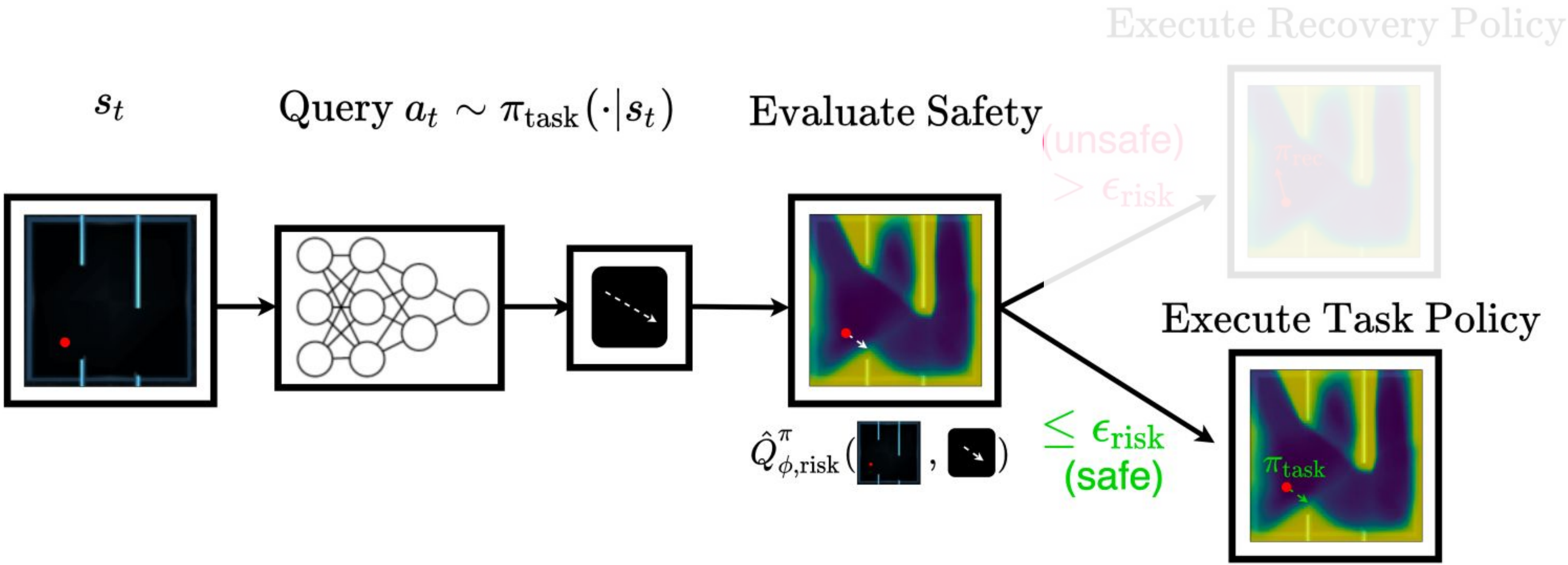
During Inference!



Thananjeyan, Brijen, et al. "Recovery RL: Safe reinforcement learning with learned recovery zones", 2021.

Anatomy of the CIMRL Model: Recovery RL

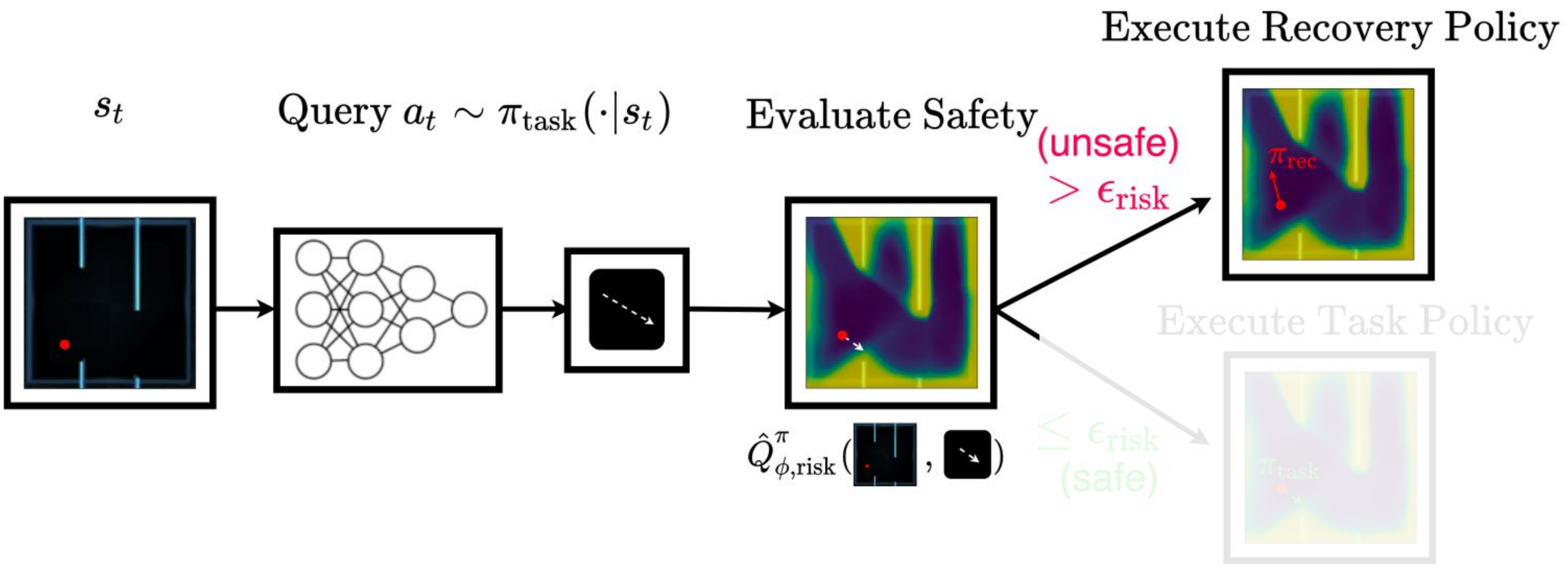
During Inference!



Thananjeyan, Brijen, et al. "Recovery RL: Safe reinforcement learning with learned recovery zones", 2021.

Anatomy of the CIMRL Model: Recovery RL

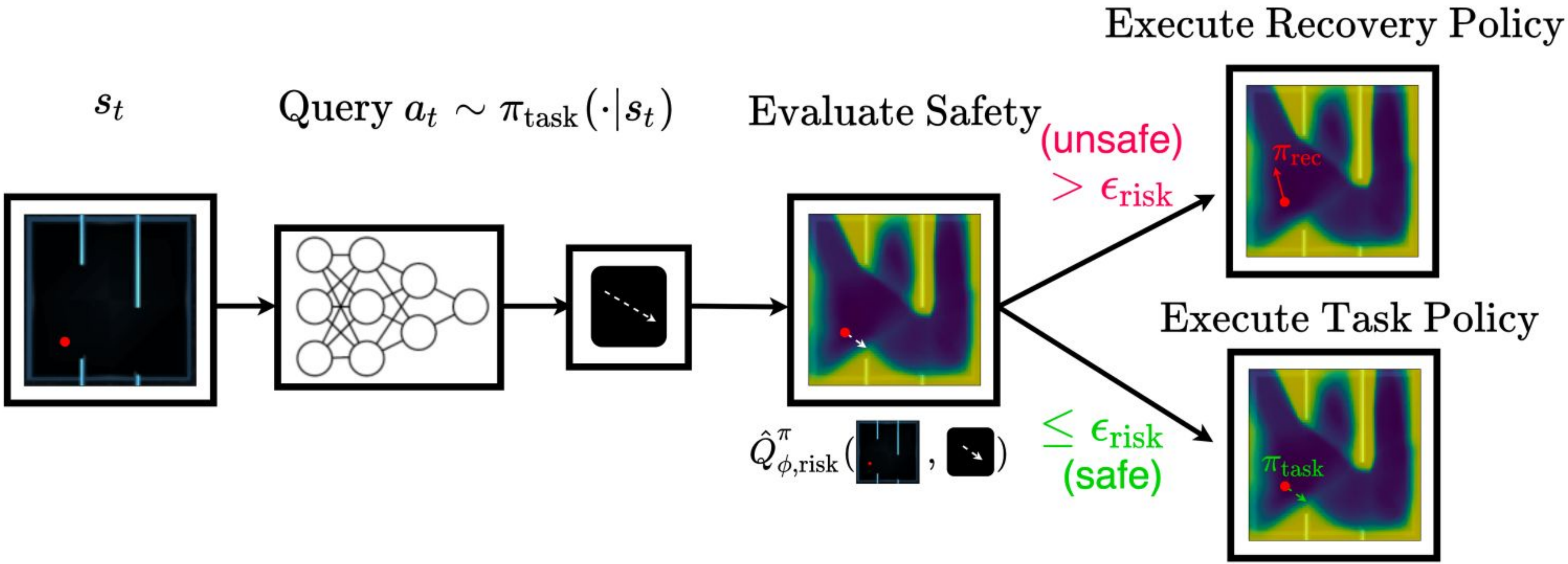
During Inference!



Thananjeyan, Brijen, et al. "Recovery RL: Safe reinforcement learning with learned recovery zones", 2021.

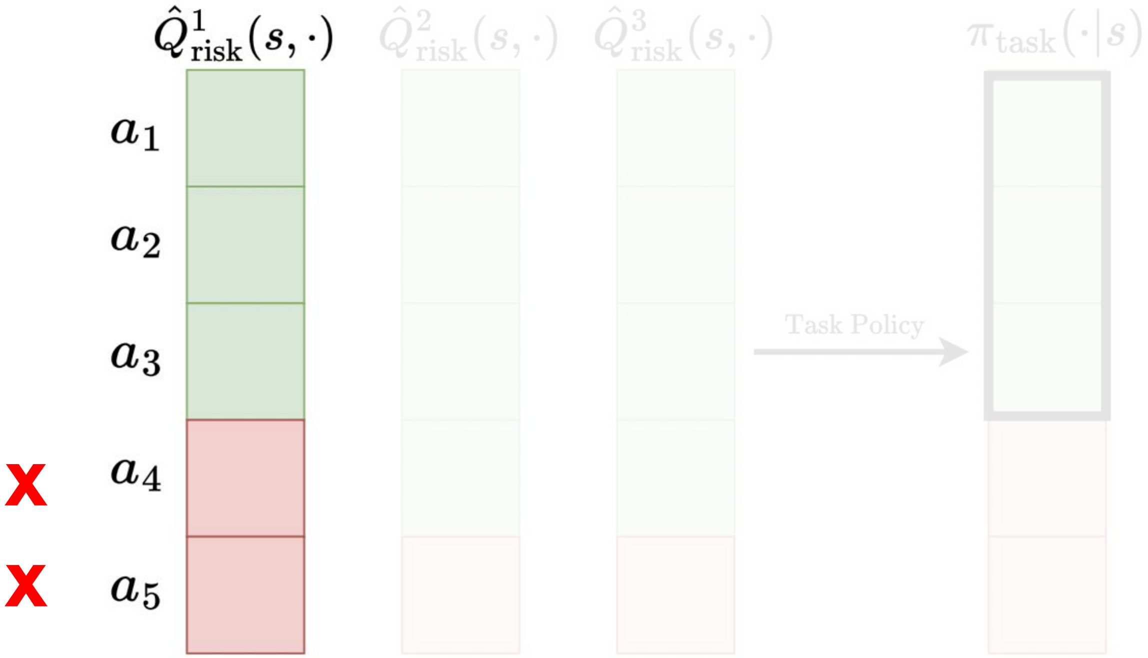
Anatomy of the CIMRL Model: Recovery RL

During Inference!

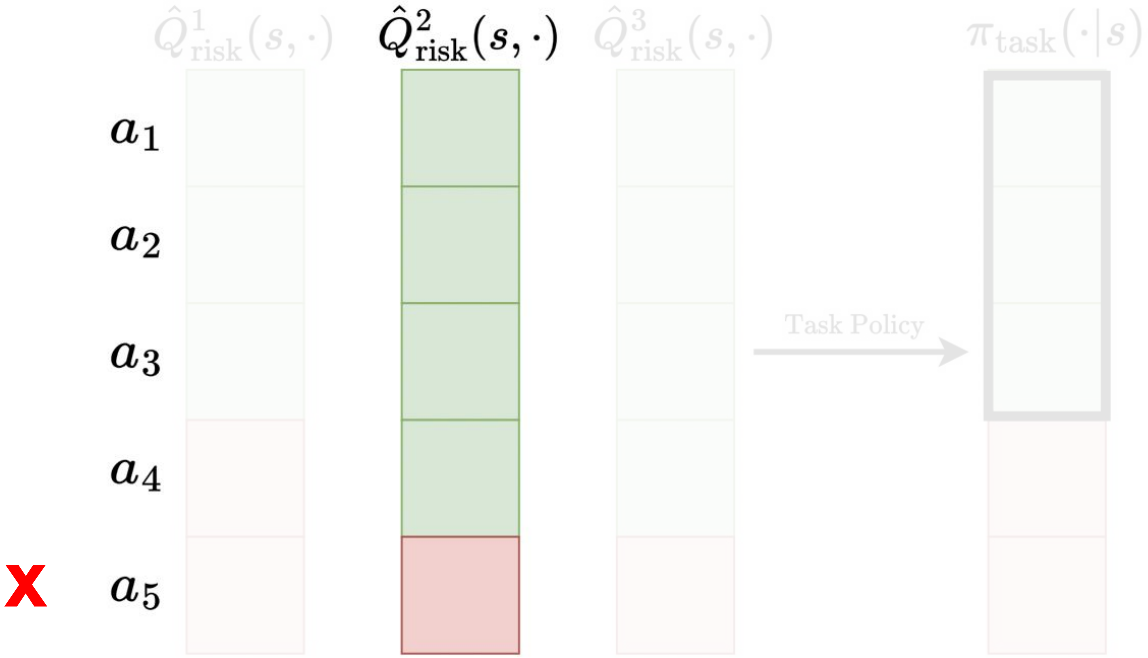


Thananjeyan, Brijen, et al. "Recovery RL: Safe reinforcement learning with learned recovery zones", 2021.

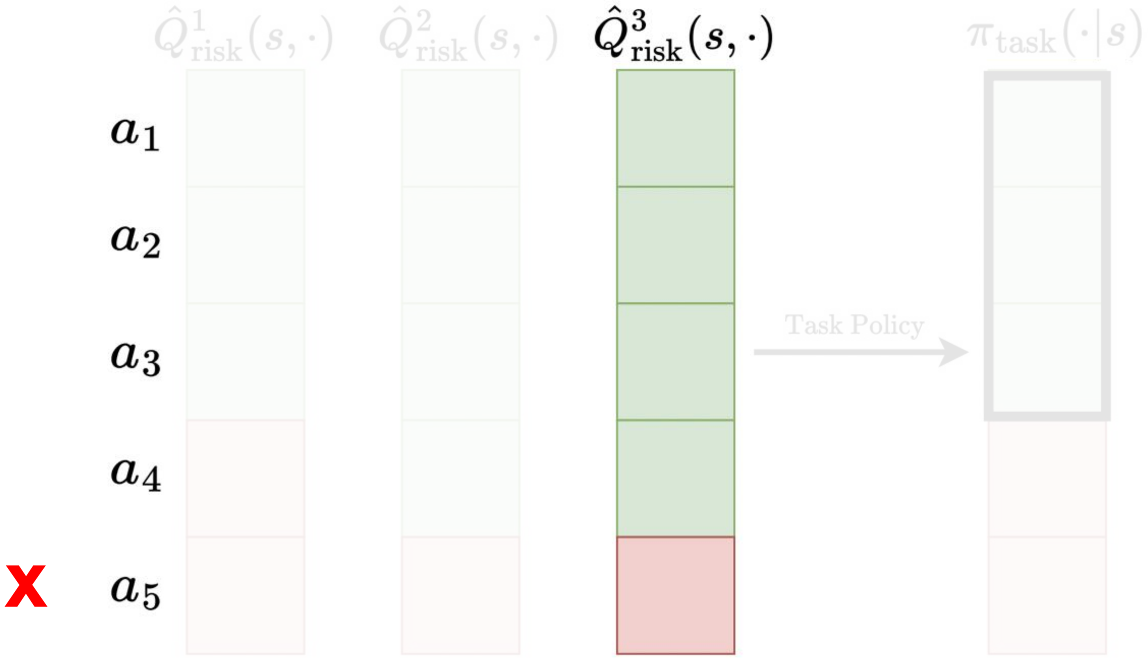
Constructing CIMRL Mixed Policy: Safe Case



Constructing CIMRL Mixed Policy: Safe Case

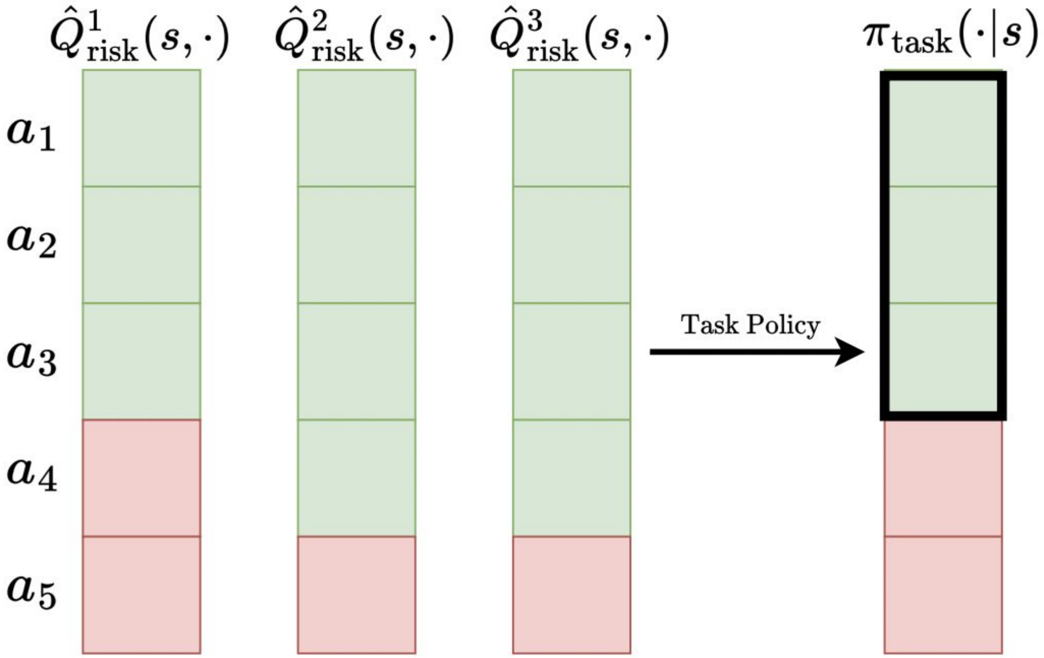


Constructing CIMRL Mixed Policy: Safe Case

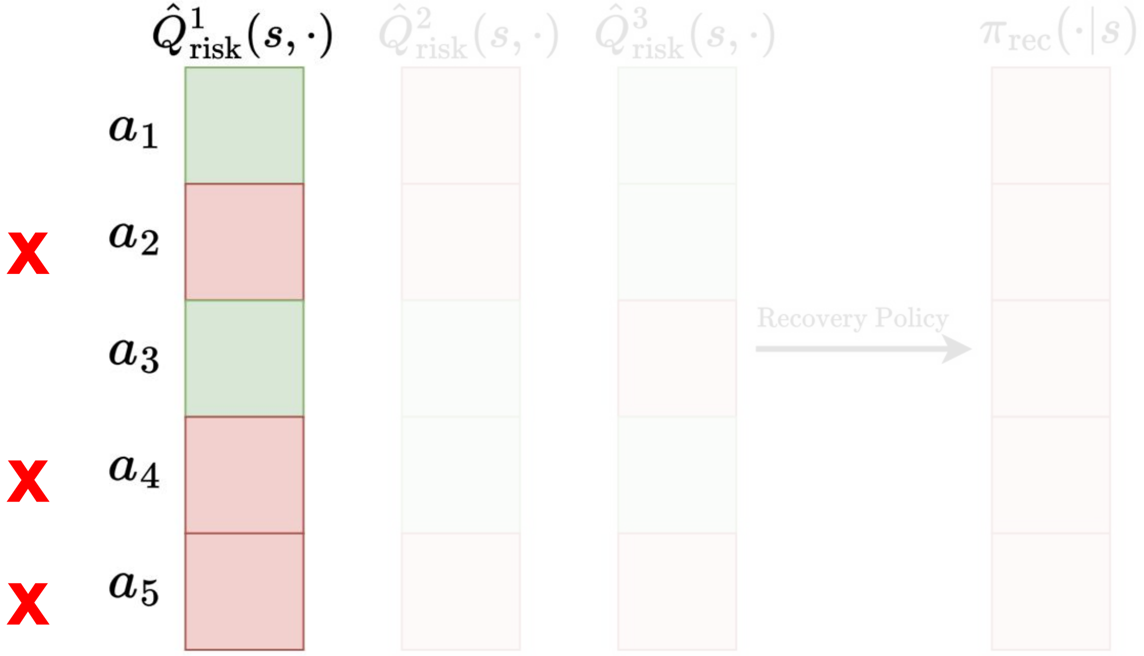


Constructing CIMRL Mixed Policy: Safe Case

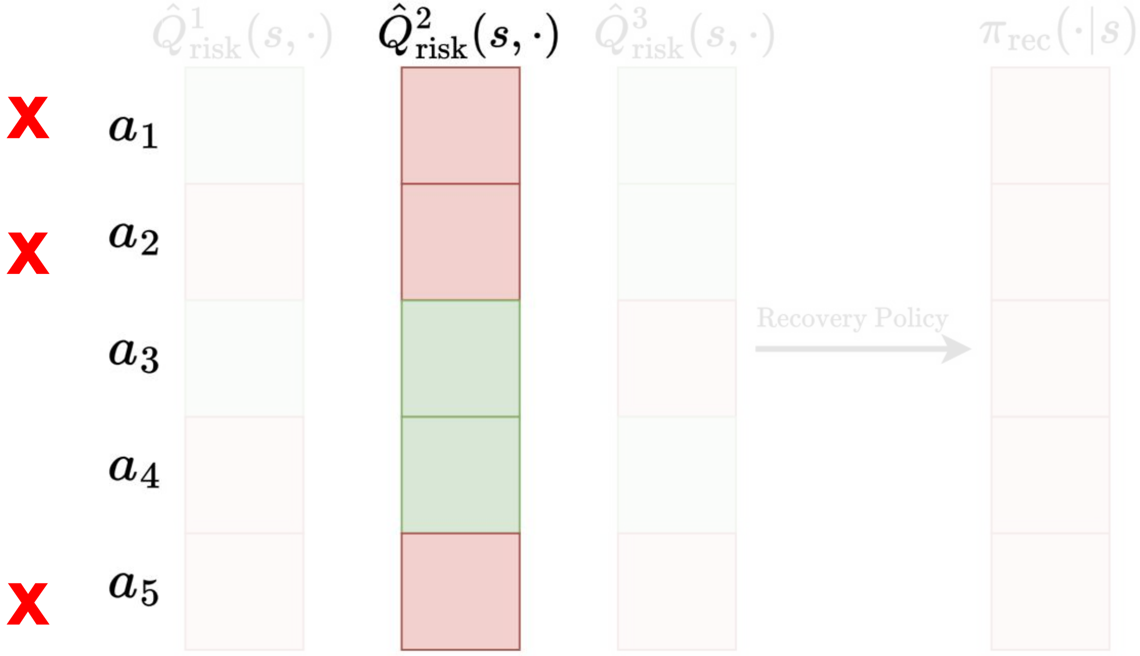
If there exist safe actions then sample from re-normalized task policy.



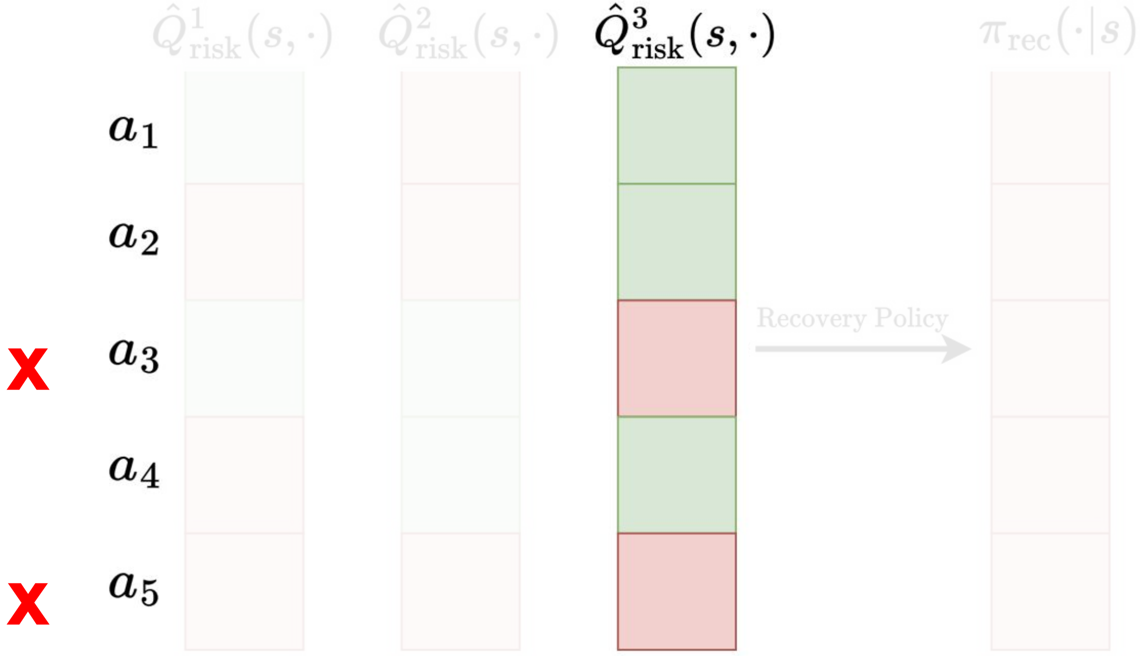
Constructing CIMRL Mixed Policy: Unsafe Case



Constructing CIMRL Mixed Policy: Unsafe Case

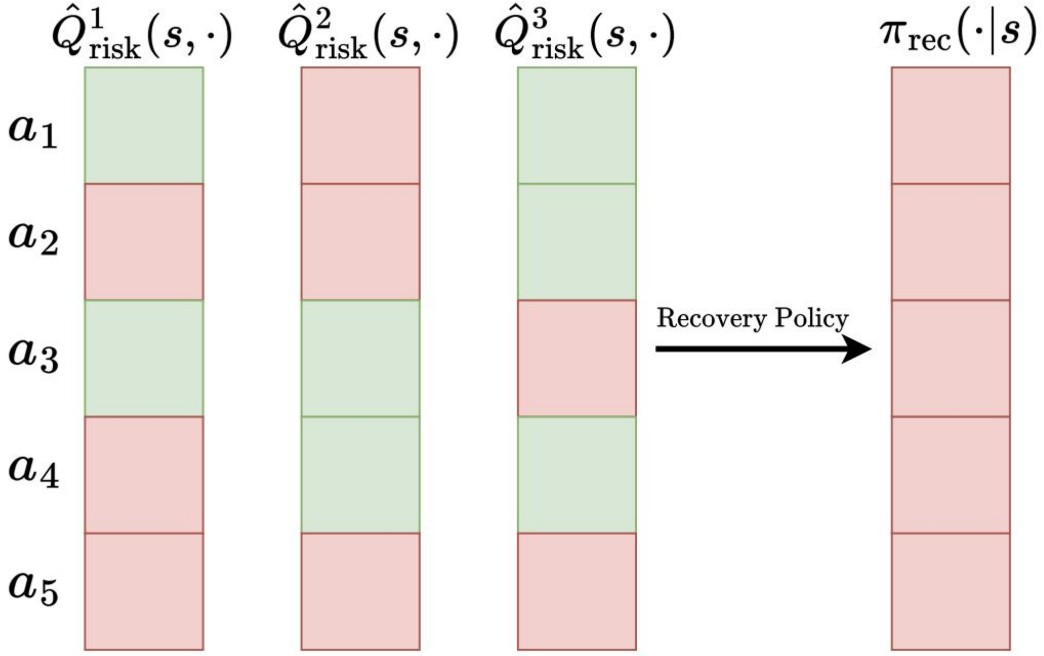


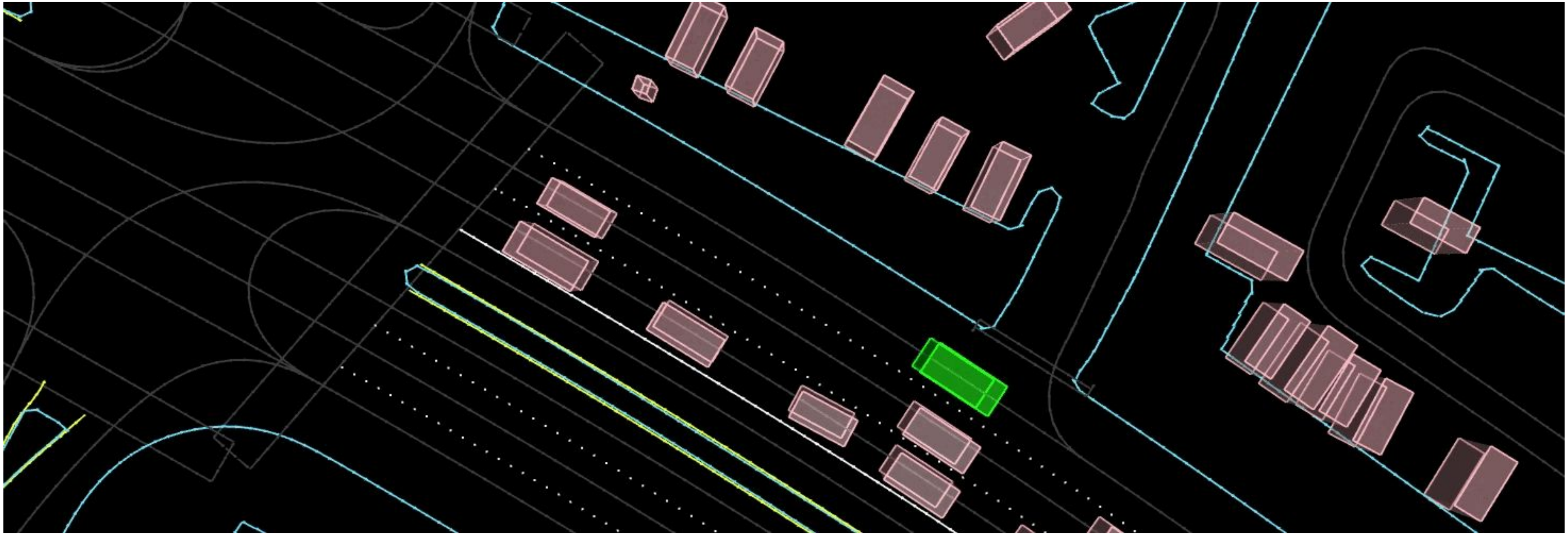
Constructing CIMRL Mixed Policy: Unsafe Case



Constructing CIMRL Mixed Policy: Unsafe Case

Otherwise sample from recovery policy





Closed-Loop Simulator

Waymax:

- ➔ Can be used for training
- ➔ Data-driven
- ➔ TPU / GPU support

https://waymo.com/research/waymax/

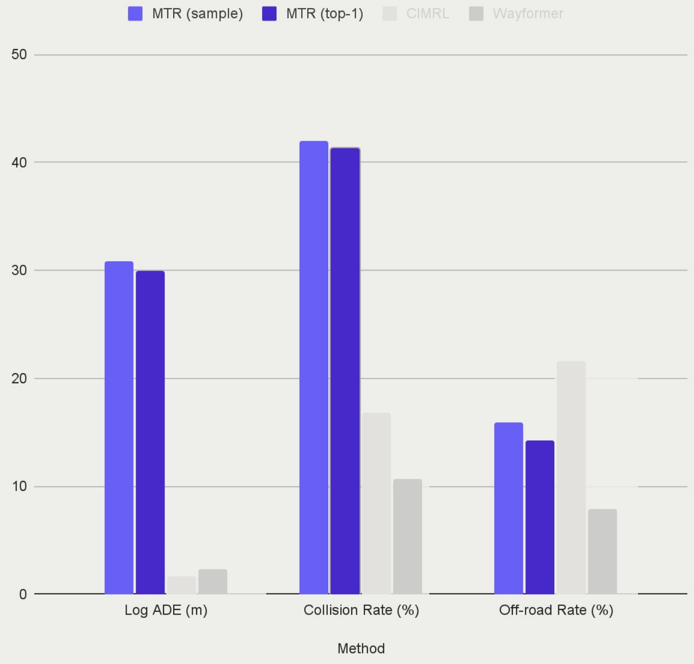
Gulino, Cole, et al. "Waymax: An accelerated, data-driven simulator for large-scale autonomous driving research." 2023.

Closed-Loop Results: Waymax

➔ Kinematic Feasibility: pretty meaningless for any Prediction-based method

➔ Route progress ratio: do not have the access to route info (*sdc_path*)

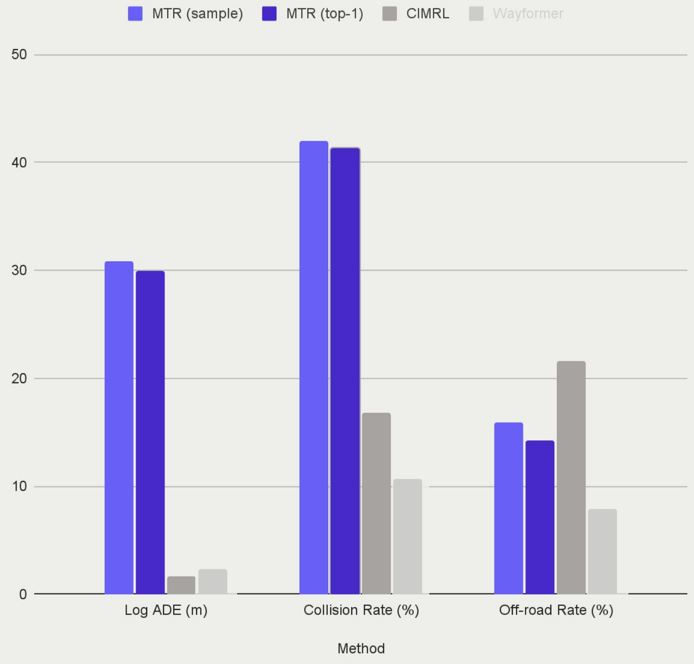
Using Waymax: No Sim Agents, Delta Action Space



Closed-Loop Results: Waymax

- ➔ Kinematic Feasibility: pretty meaningless for any Prediction-based method
- ➔ Route progress ratio: do not have the access to route info (*sdc_path*)

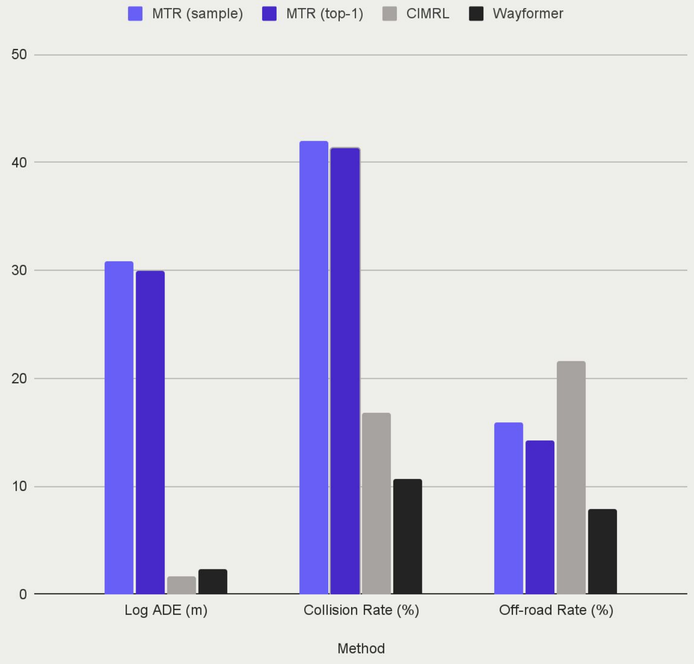
Using Waymax: No Sim Agents, Delta Action Space



Closed-Loop Results: Waymax

Wayformer has the access to route info :)

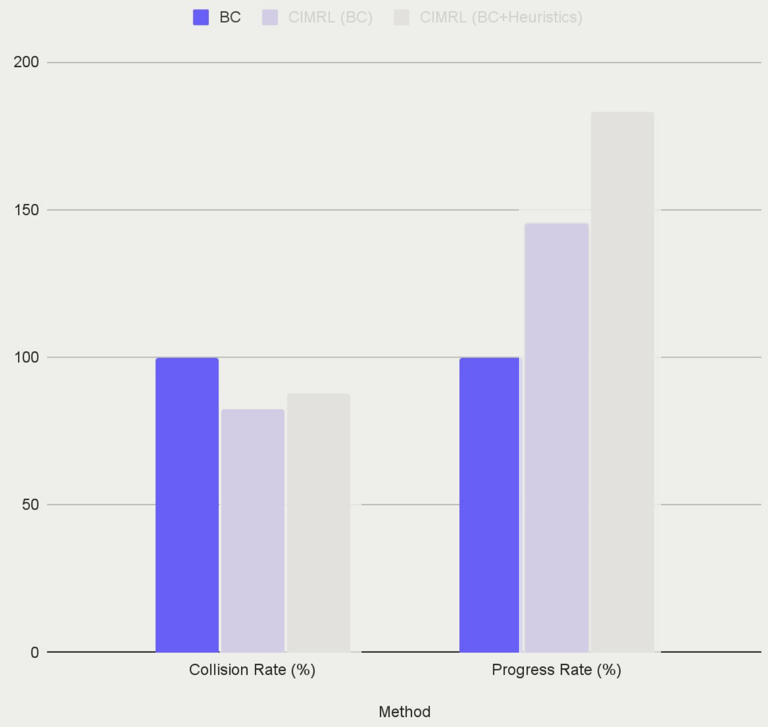
Using Waymax: No Sim Agents, Delta Action Space



Closed-Loop Results: In-house

- ➔ Challenging interactive in-house scenes where log pose divergence is usually inevitable
- ➔ Route progress ratio: makes sense
- ➔ Log ADE: doesn't

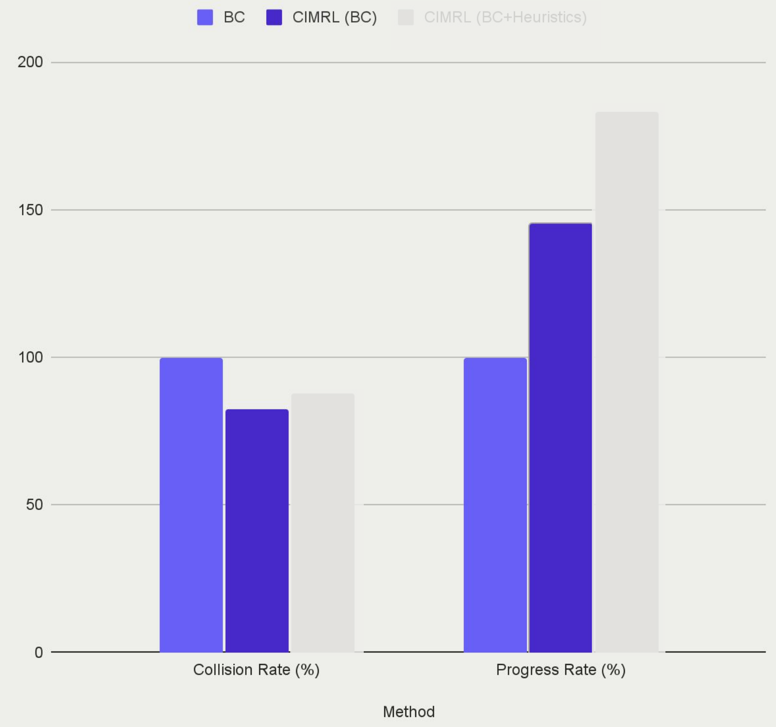
Using Internal data and Sim (Log replay)



Closed-Loop Results: In-house

- ➔ Challenging interactive in-house scenes where log pose divergence is usually inevitable
- ➔ Route progress ratio: makes sense
- ➔ Log ADE: doesn't

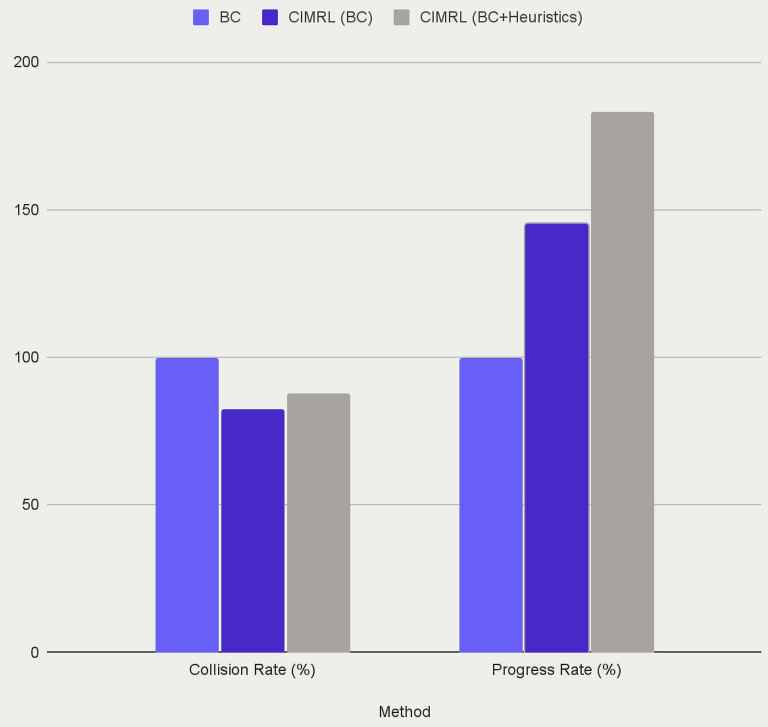
Using Internal data and Sim (Log replay)



Closed-Loop Results: In-house

- ➔ Challenging interactive in-house scenes where log pose divergence is usually inevitable
- ➔ Route progress ratio: makes sense
- ➔ Log ADE: doesn't

Using Internal data and Sim (Log replay)



CIMRL: Limitations

… And still dependent on the quality of the underlying ego plan generation procedure.

01

Reward definition is not straightforward (but *mitigatable*)

02

Rare sparse events are challenging to learn (i.e. *collisions*) esp. for advanced planners

03

Sample inefficient – takes many simulation steps to learn (*huge* state-action space)

Conclusions

01

CIMRL is really scalable and flexible framework of combining paradigms

02

Learning selection provides long-horizon reasoning

03

There is no such a thing as “too much safety” :(

New Horizons

New RnD direction in FinTech opens now!

If you feel comfortable to *understand*, *implement*, and *push forward* the Tech inside Finance - contact me with your CV!

https://petiushko.info/#contact

Thanks!