11/18/2025

Safe Planning

In Autonomous Driving

Aleksandr Petiushko, https://petiushko.info
Professor @ Sofia University

Contributors

Ashwin Balakrishna

Jonathan Booher

Ishan Gupta

Vladislav Isenbaev

Bo Li

Wei Liu

Aleksandr Petiushko

Khashayar Rohanimanesh

Taiqi Wang

Junhong Xu

Xuan Yang

Yu Yao

Jiawei Zhang

*Mentioned in the alphabetical order

Autonomous Driving (AD)

- ***AD*** is one of the most complex and difficult tasks, both theoretically and practically
- ***Planning*** is a key focus regarding ***safety***

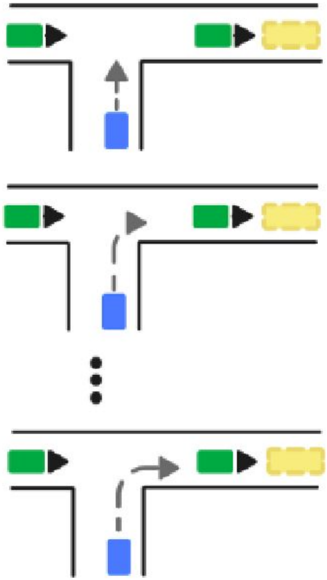
Image source

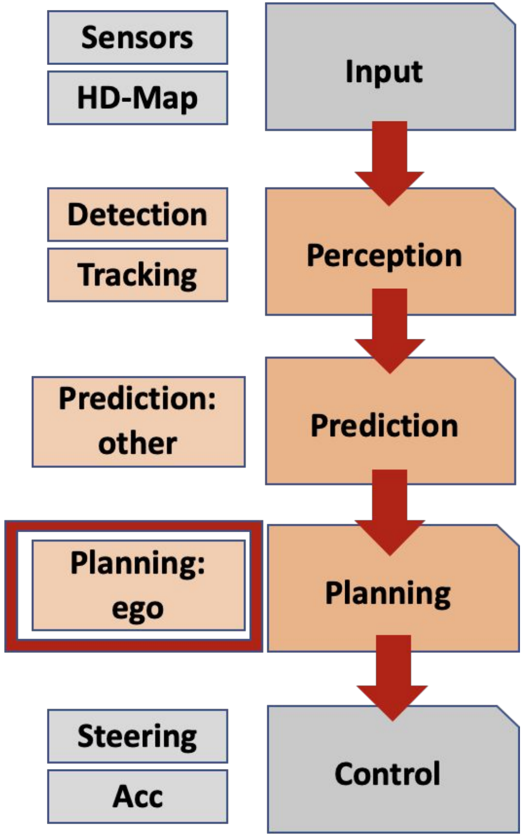
Safety of AVs on the road is crucial

How to choose / check the right plan?

- Add a violation **checker**!

- Need a **scorer**!

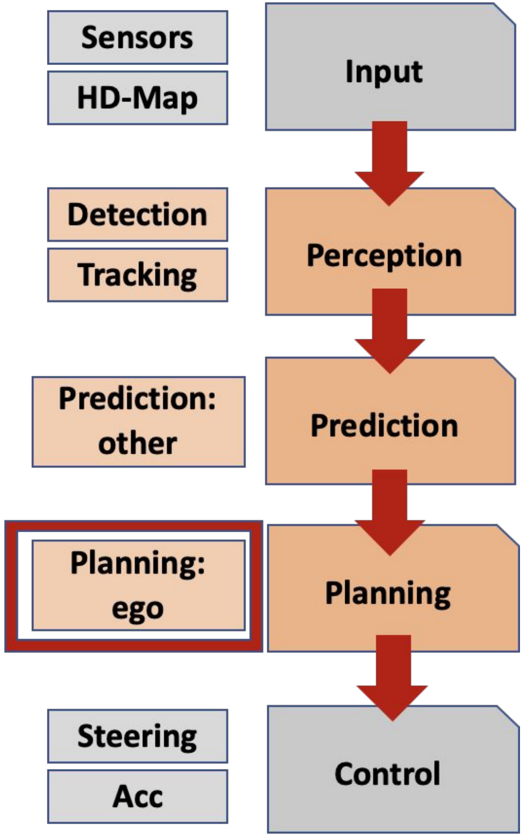




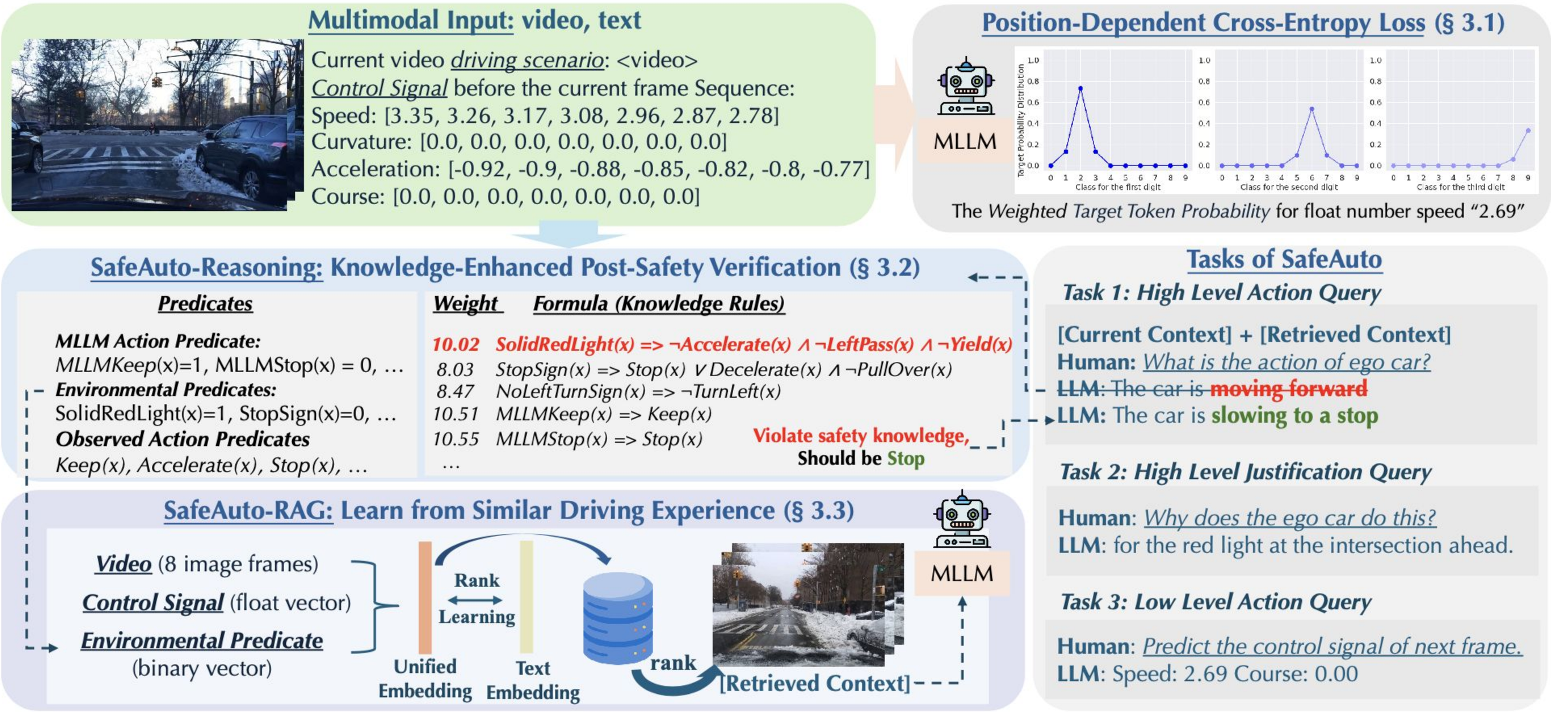
How to choose / check the right plan?

- Add a violation **checker**!
- Need a **scorer**!

?



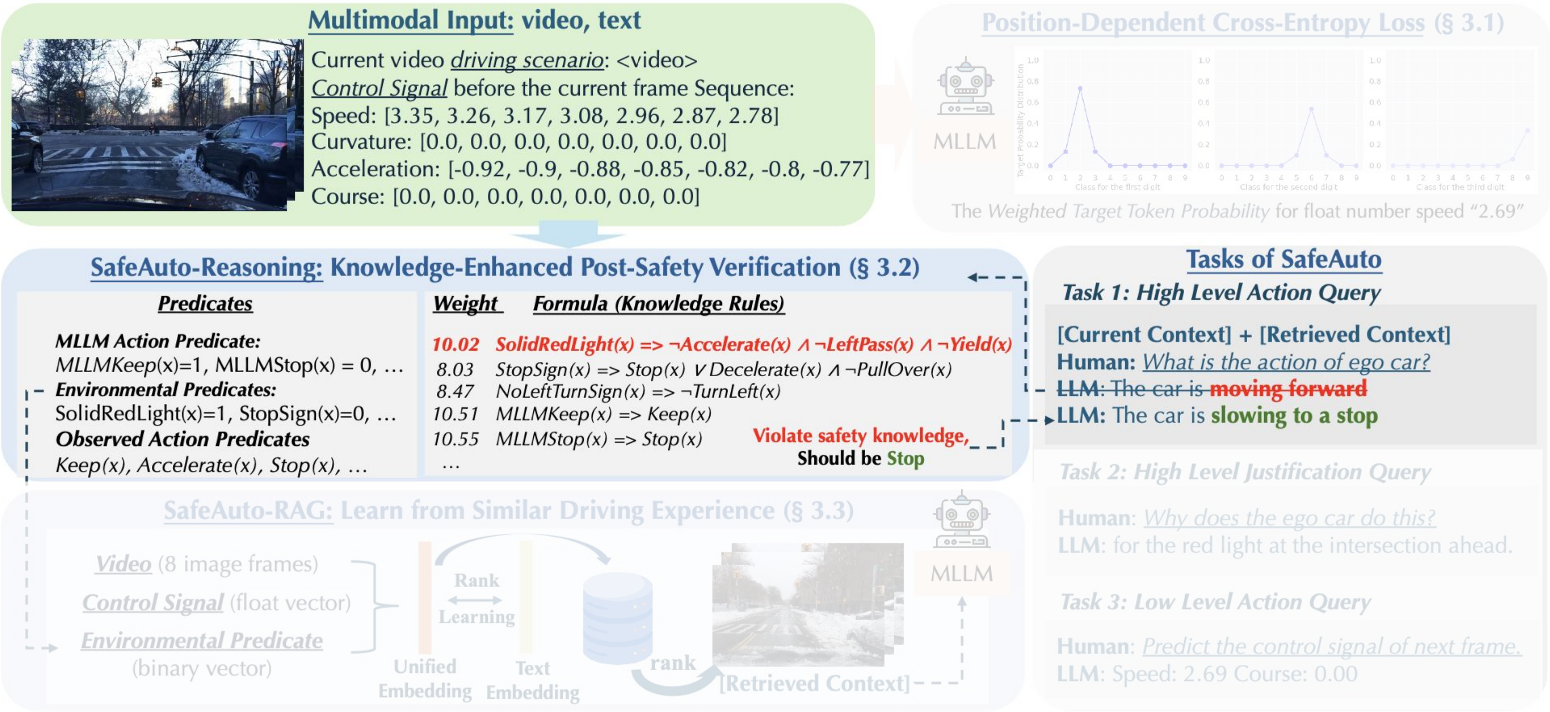
SafeAuto: the Overall Approach[1]



***MLLM** = multimodal large language model*

[1] Zhang, Jiawei, et al. "[SafeAuto](https://openreview.net/forum?id=nKJGjovmZz): Knowledge-Enhanced Safe Autonomous Driving with Multimodal Foundation Models". ICML 2025.
https://openreview.net/forum?id=nKJGjovmZz

SafeAuto: the Overall Approach[1]



MLLM = *multimodal large language model*

[1] Zhang, Jiawei, et al. "SafeAuto: Knowledge-Enhanced Safe Autonomous Driving with Multimodal Foundation Models". ICML 2025. https://openreview.net/forum?id=nKJGjovmZz

Markov Logic Networks

- ➔ Currently, most MLLMs are still *data-driven*
- ➔ Reliability and strict adherence to safety regulations are inevitable
- ➔ Let's use *Probabilistic Graphical Models* to verify the safety
    - **Markov Logic Networks** (MLN) to combine:
        - *Domain knowledge*
        - *Traffic rules*

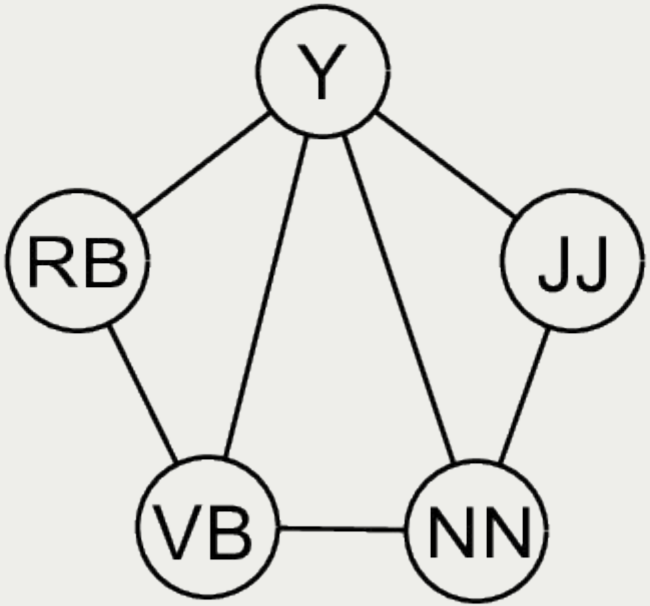


Image source

MLNs Details

- MLN == a set of *first-order logic formulas* with an associated confidence *weight **w***
  - ***w***: to model uncertainty / deal with exceptions in real-world knowledge
  - Ex.: a traffic rule like "***If there is a stop sign, then the vehicle should stop or decelerate***" can be represented as the logical formula:
    - `StopSign(x)` ⇒ `Stop(x)` ∨ `Decelerate(x)`

$$P(\mathrm{X}) = \frac{1}{Z}\exp\left(\sum_{f\in F}\omega_f \sum_{a_f\in A_f}\phi_f(a_f)\right)$$

where:

- **X**: set of all ground truth predicates
- *Z*: partition function
- $\phi_f(a_f)$: potential function for formula *f* with assignment a_f (=1 iff a_f)
- *F*: set of all formulas *f*
- A_f: set of all possible assignments to the arguments of formula *f*

MLN in AD

- Predicates:
  - **Unobserved** *U*:
    - Vehicle should take (*Stop*, *Accelerate*, *TurnLeft*)
  - **Observed** *O*:
    - MLLM Action (*MLLMStop*, *MLLMAccelerate*, *MLLMTurnLeft*)
      - MLLMStop => Stop
    - Environmental (*StopSign*, *SolidRedLight*)
      - From video, using YOLOv8 [1] trained on LISA [2]
      - + Historical Control Signal (*HCSTurnLeft*)

$$\text{StopSign}(x) \Rightarrow$$

$$\Rightarrow \text{Stop}(x) \vee \text{Decelerate}(x) \wedge \neg\text{PullOver}(x)$$

Example of environmental *observed* predicate

[1] YOLO model
[2] LISA dataset

MLN in AD - Process (1)

➔ **Inference**
- Obtain the most realistic *unobservable U* given the *observable O* using the trained *MLN*

$$\mathcal{U}^* = \arg\max_{\mathcal{U}} P(\mathcal{U}|\mathcal{O})$$

➔ **Training**
- Obtain the *weights* $\boldsymbol{w_f}$ to maximize the $P(U|O)$ with BDD-X [1] / DriveLM [2] data

$$\omega_f$$

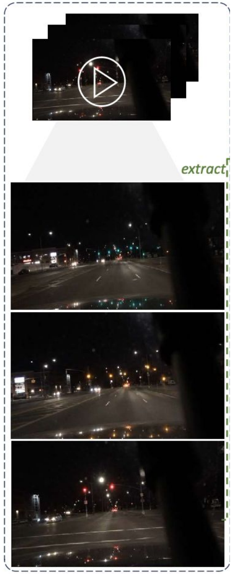
➔ **Safety verification**
- After inferring the *U* based on *O* from *MLN*, ***if*** it contradicts MLLM's action (a potential *safety violation* / *breach* of critical *traffic rules*) **=>** need to overwrite the high-level action query and *re-prompt* the MLLM *again*

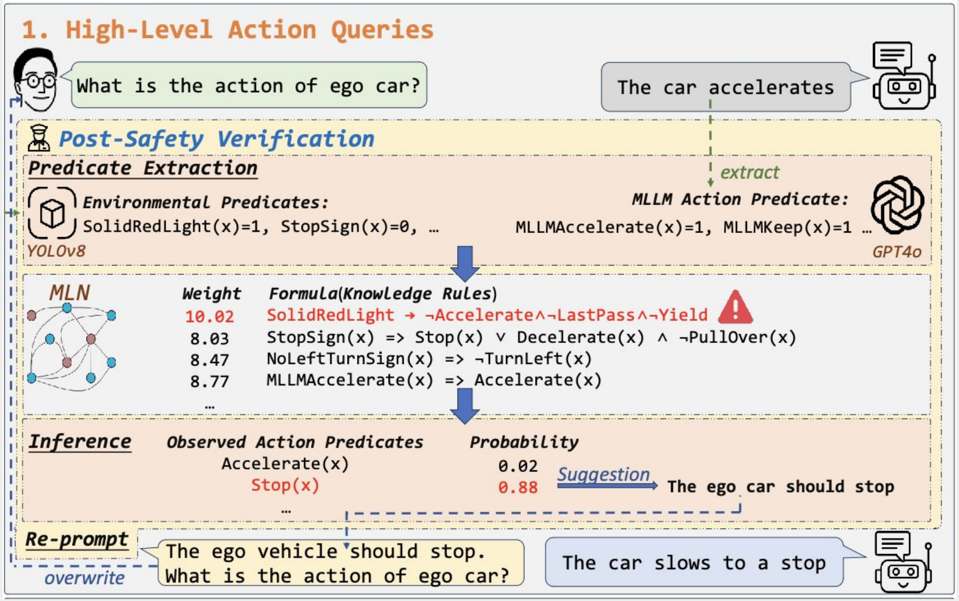
[1] BDD-X dataset
[2] DriveLM dataset

MLN in AD - Process (2)

➔ **MLN**

- Serves as a ***post-verification layer*** able to change the unsafe MLLM system initial suggestion
- **Improving** the overall **trust** to AD system





MLN in AD - Results

➔ Ablation study on the **impact** of each module on the traffic rule violation rate of MLLM-predicted actions

Method	BDD-X	DriveLM
Base	11.64%	1.03%
PDCE	8.44%	1.46%
PDCE+RAG	5.90%	1.03%
PDCE+RAG+MLN	***4.50%***	***0.75%***

(lower the better)

DriveLM use case

Method	High-Level Behavior			Motion
	Accuracy	Speed	Steer	ADE
Base	60.58	64.57	80.29	0.86
PDCE	63.21	67.88	79.27	0.85
PDCE+MLN	***66.86***	***71.39***	***80.29***	***0.85***
PDCE+RAG	74.01	79.27	81.61	0.84
PDCE+RAG+MLN	***74.61***	***79.85***	***81.91***	***0.84***

(higher the better) *(lower the better)*

MLN in AD: Outcomes

- **Markov Logic Network** provides an additional layer of safety in AD
- Limitations:
  - Need to understand the **Markov**-based reasoning
  - Doesn't work **equally** best for every dataset
  - **One more** ML model

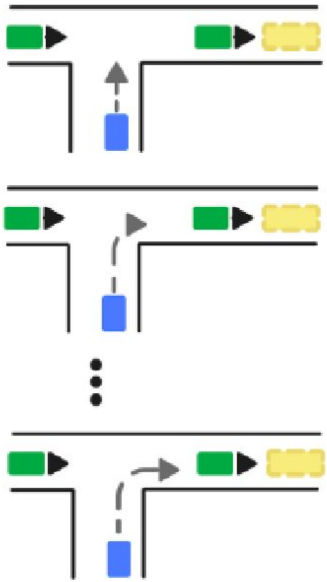
BDD-X use case

Method \ Metric	Action / Meteor	Action / Accuracy	Justification / Meteor
Base	29.2	61.75	13.2
PDCE	29.3	61.94	13.2
PDCE+MLN	***29.4***	***62.97***	***13.2***
PDCE+RAG	35.3	91.00	13.9
PDCE+RAG+MLN	***35.5***	***92.18***	***14.0***

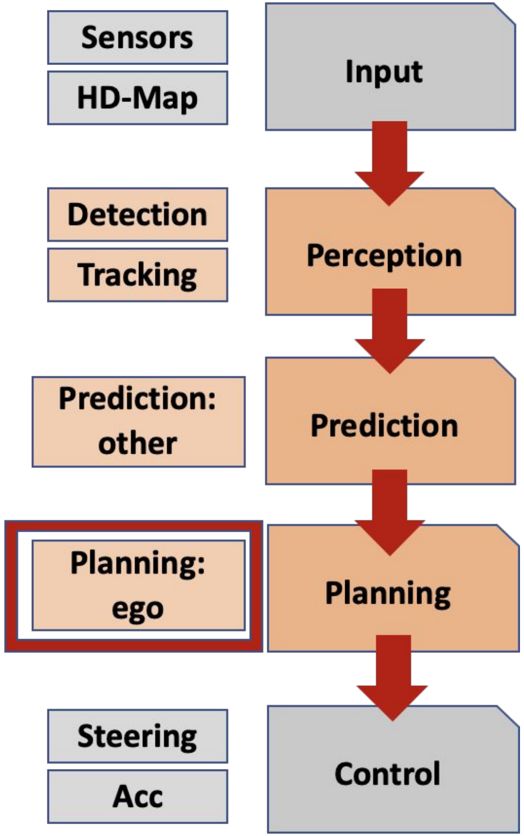
How to choose / check the right plan?

- Add a violation **checker**!

- Need a **scorer**!

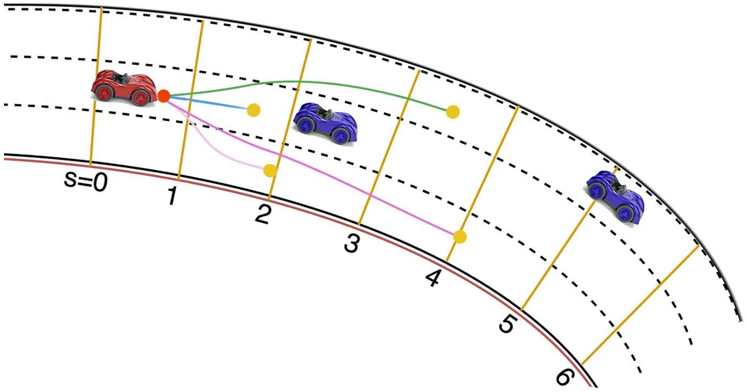


?

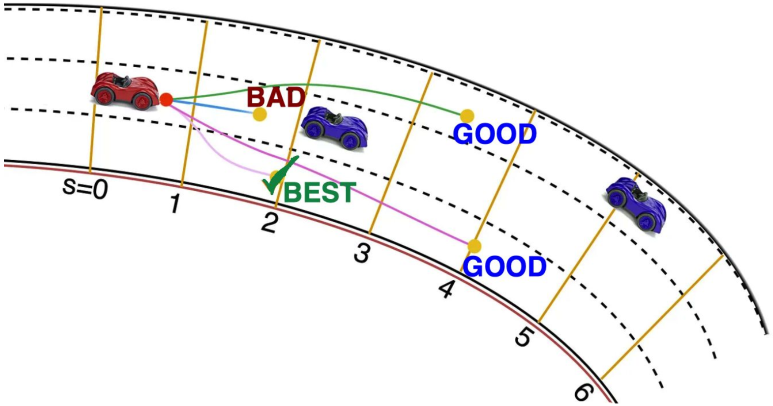


Plan Generation vs Plan *Selection*

Generation



Selection



Plan Generation vs Plan Selection (Image source)

Let's **combine** two worlds!

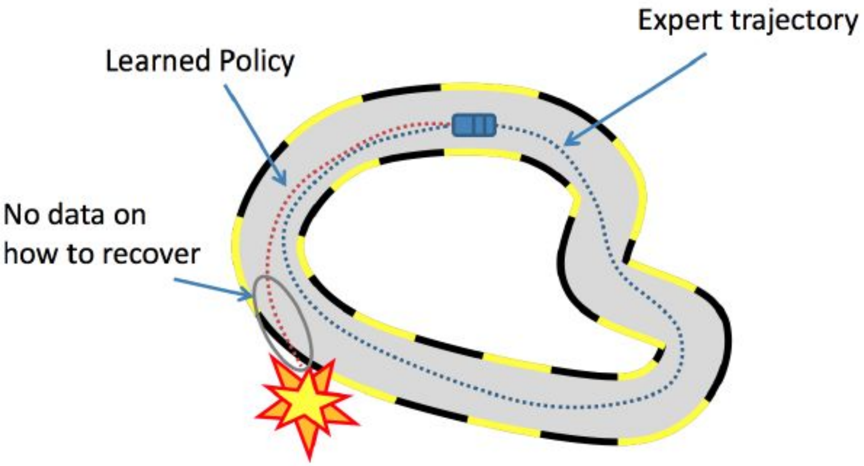
Imitation Learning

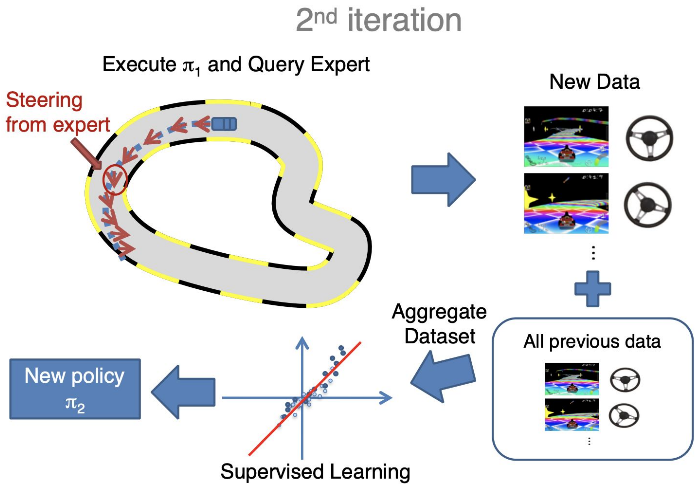
Pros:

- ➔ 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.

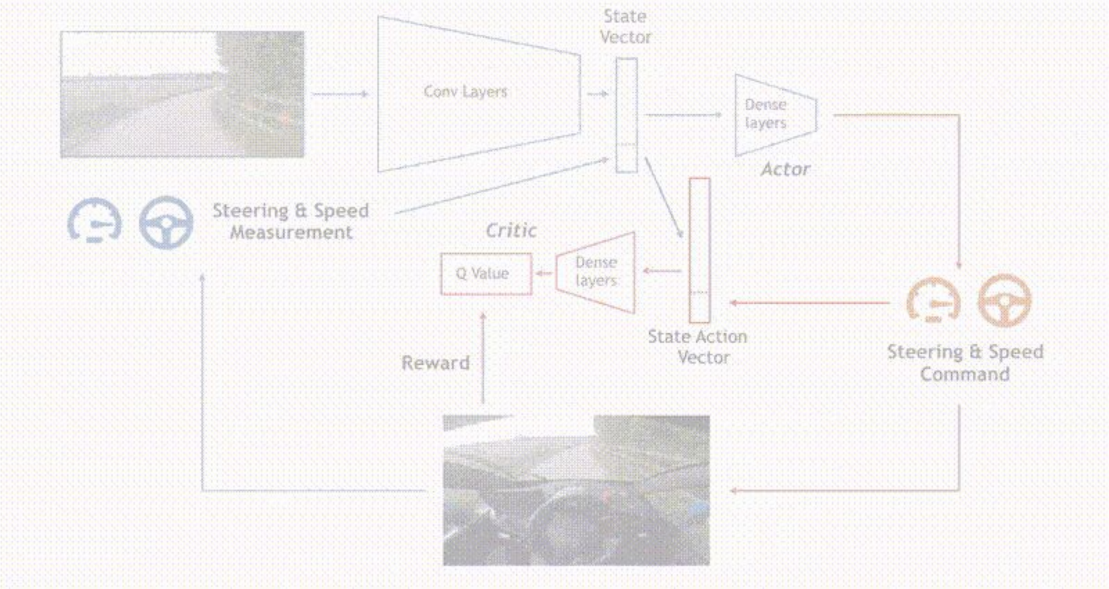
Reinforcement Learning

Pros:

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

Cons:

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



Online, off-policy RL (DDPG) from 2018

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

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

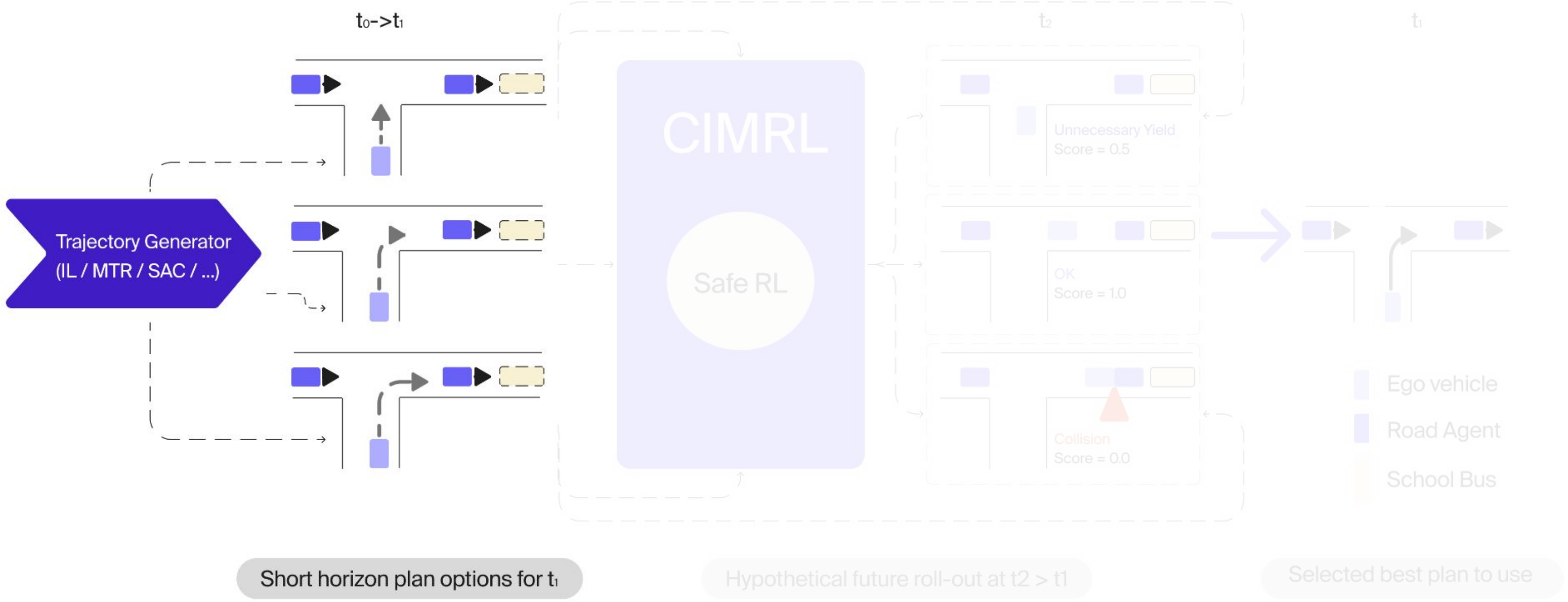
What if:

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

Plus:

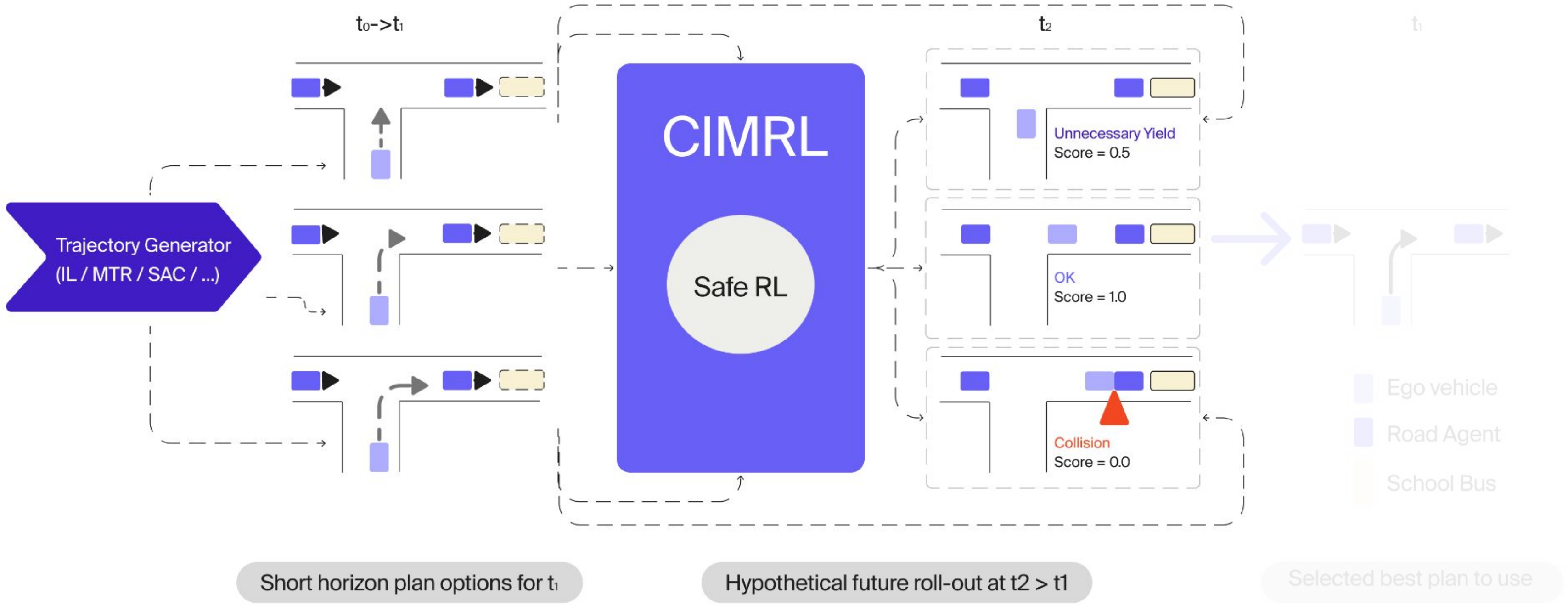
- ➔ We can concentrate on safety by doing hypothetical future roll-outs and remove / downvote dangerous plans, and provide behavior realism from IL

CIMRL[1]: Combining IMitation and Reinforcement Learning



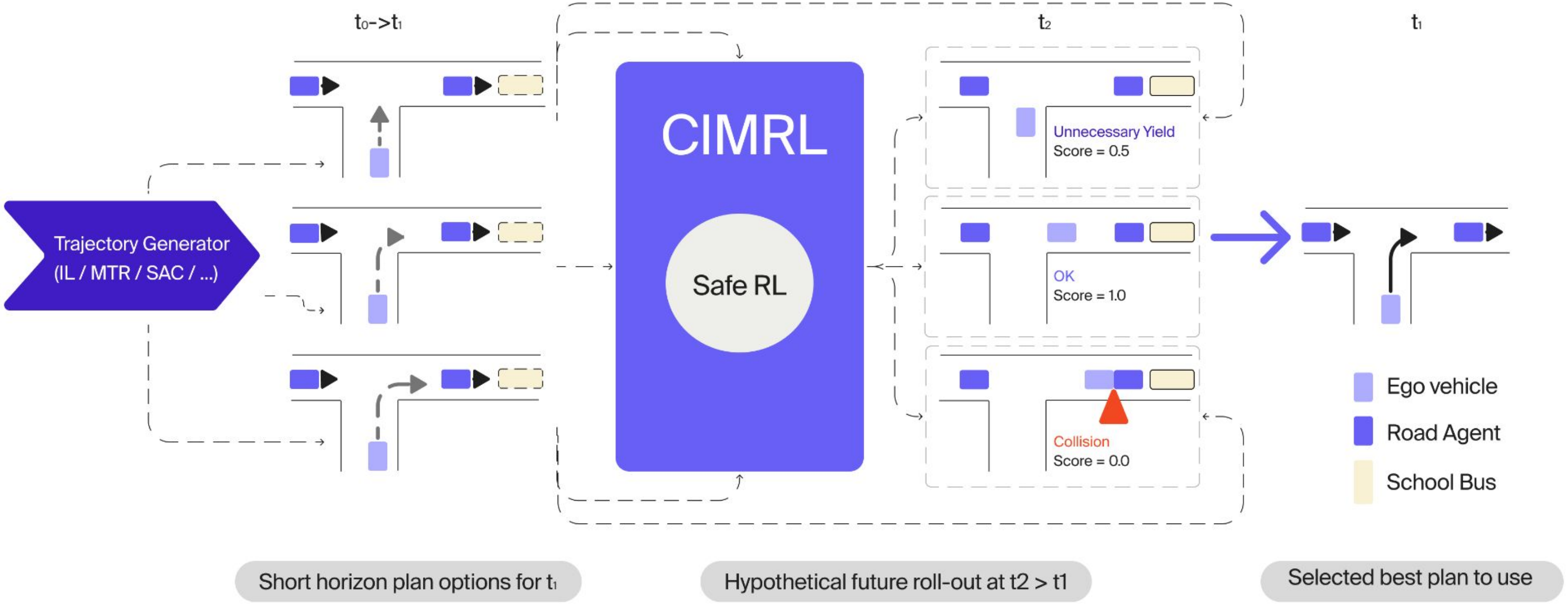
[1] Booher, Jonathan, et al. "CIMRL: Combining IMitation and Reinforcement Learning for Safe Autonomous Driving." 2024.
https://arxiv.org/abs/2406.08878

CIMRL[1]: Combining IMitation and Reinforcement Learning



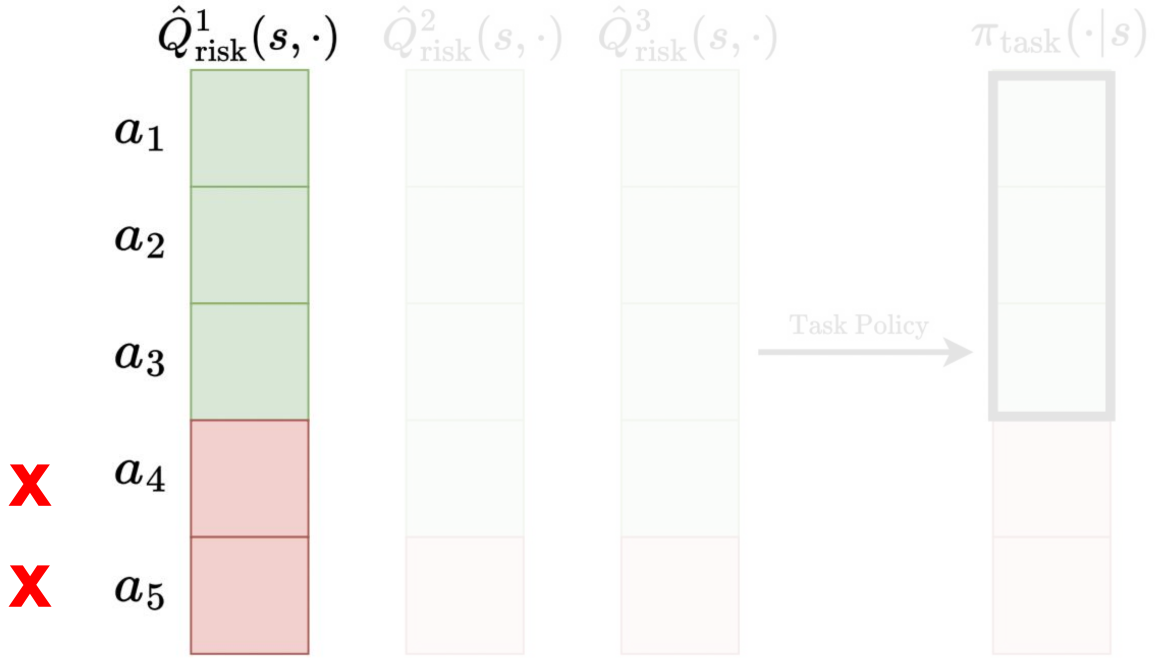
[1] Booher, Jonathan, et al. "CIMRL: Combining IMitation and Reinforcement Learning for Safe Autonomous Driving." 2024. https://arxiv.org/abs/2406.08878

CIMRL[1]: Combining IMitation and Reinforcement Learning

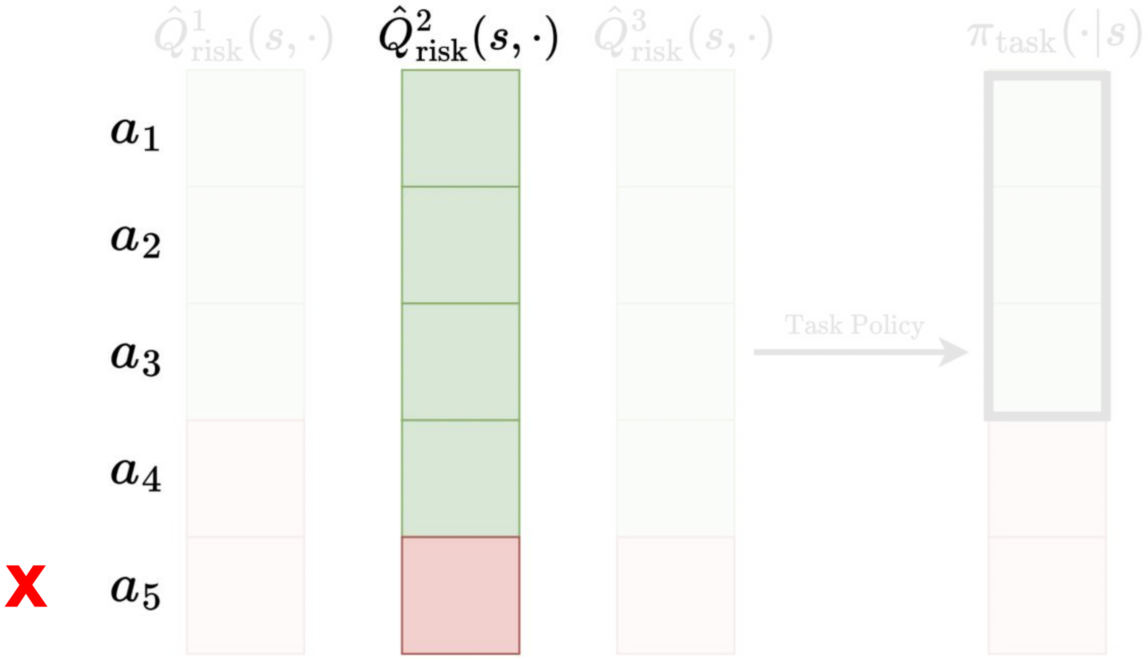


[1] Booher, Jonathan, et al. "CIMRL: Combining IMitation and Reinforcement Learning for Safe Autonomous Driving." 2024. https://arxiv.org/abs/2406.08878

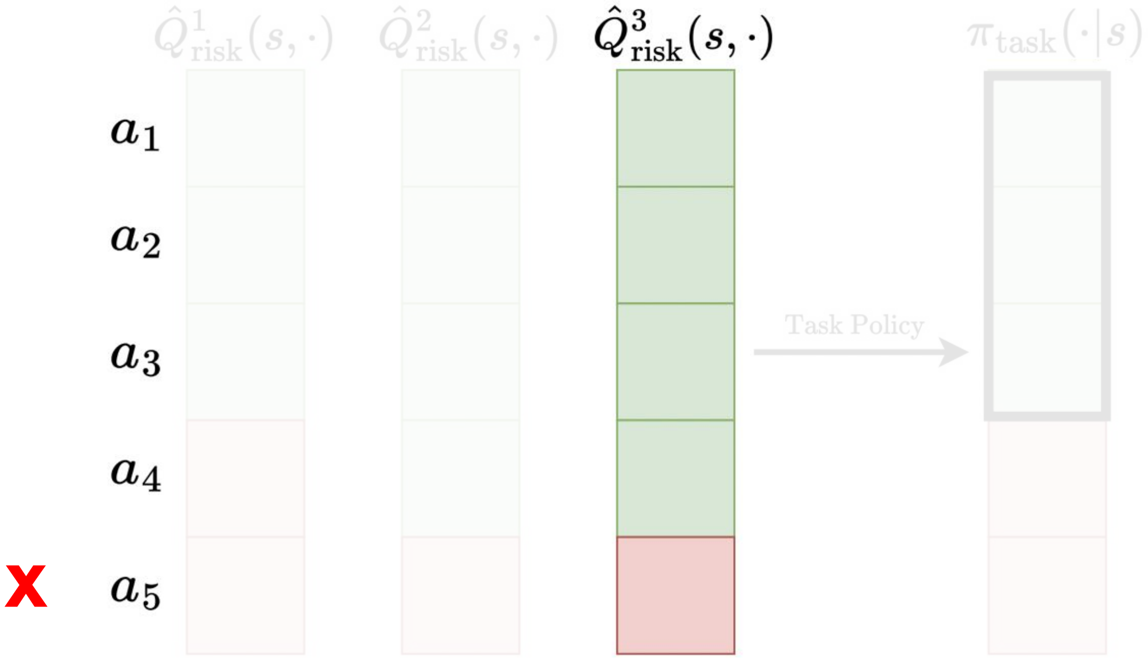
Constructing CIMRL Mixed Policy: Safe Case



Constructing CIMRL Mixed Policy: Safe Case

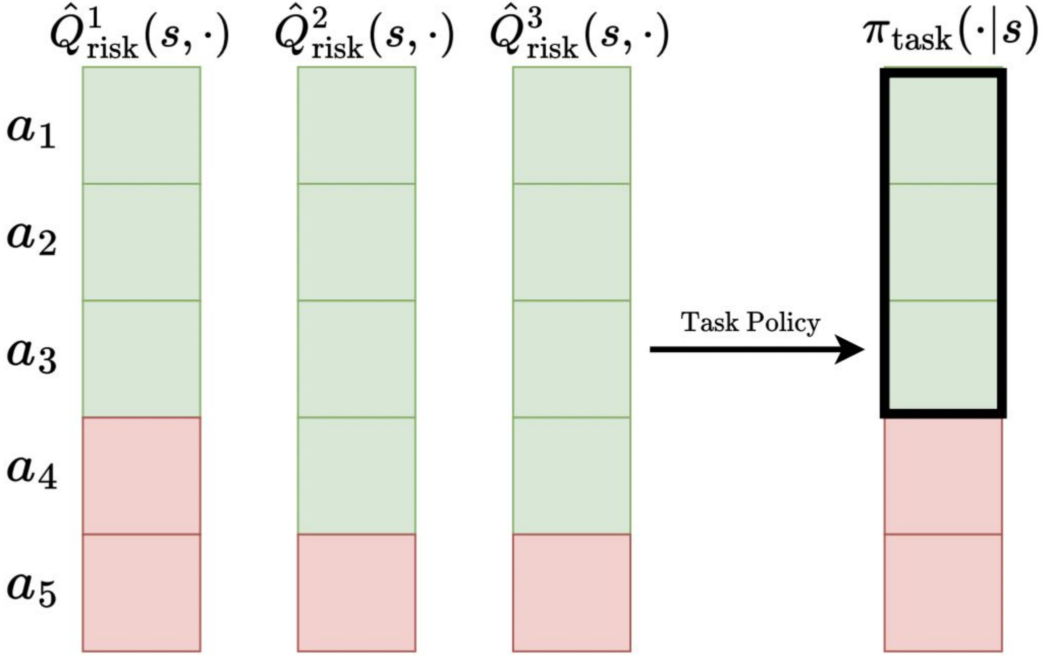


Constructing CIMRL Mixed Policy: Safe Case

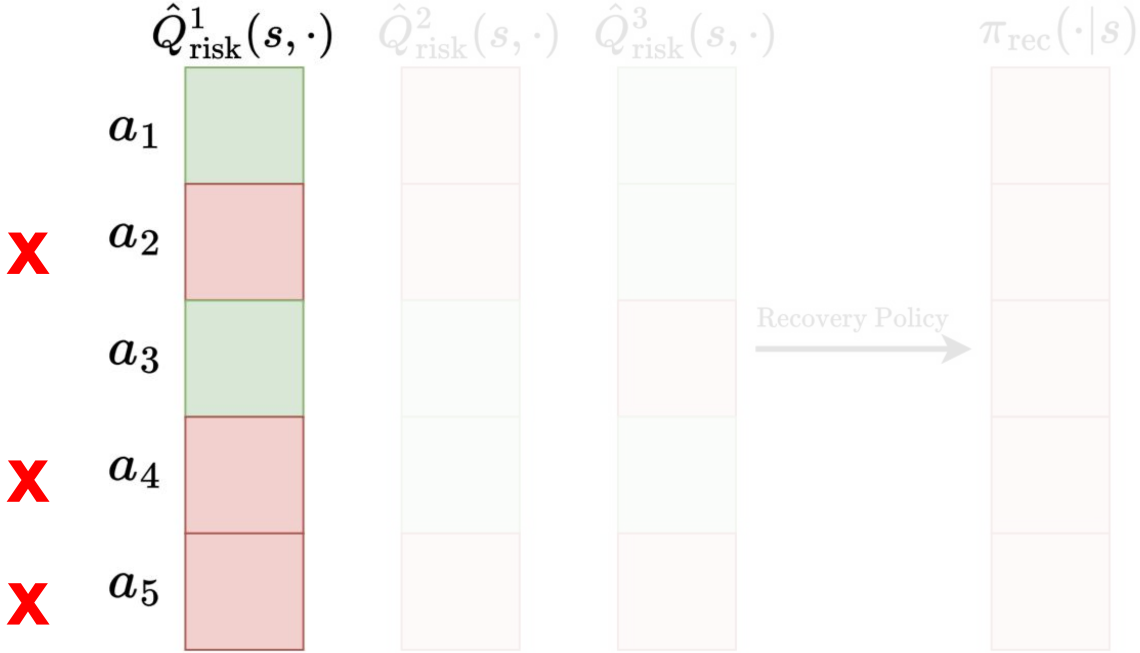


Constructing CIMRL Mixed Policy: Safe Case

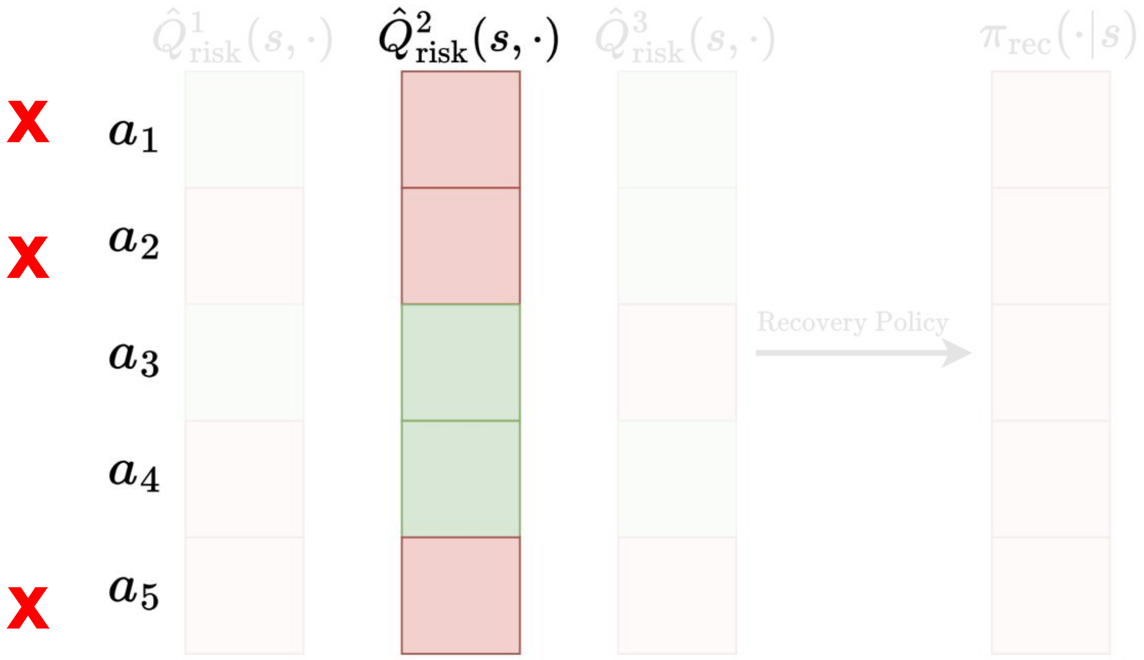
If there exist safe actions then sample from re-normalized task policy.



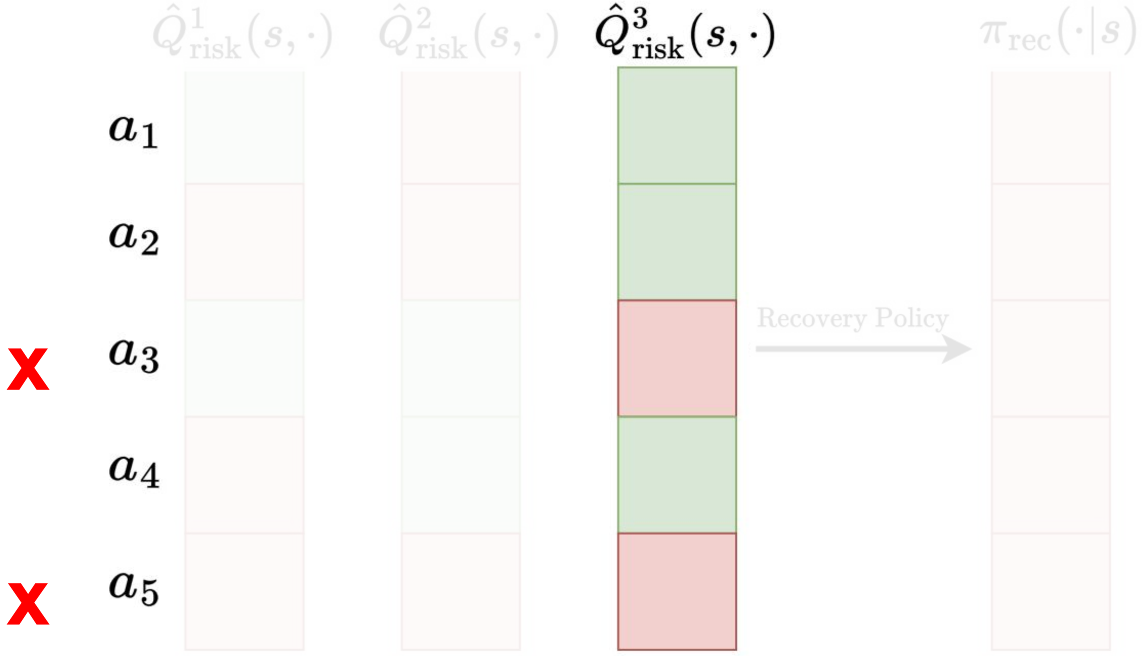
Constructing CIMRL Mixed Policy: Unsafe Case



Constructing CIMRL Mixed Policy: Unsafe Case

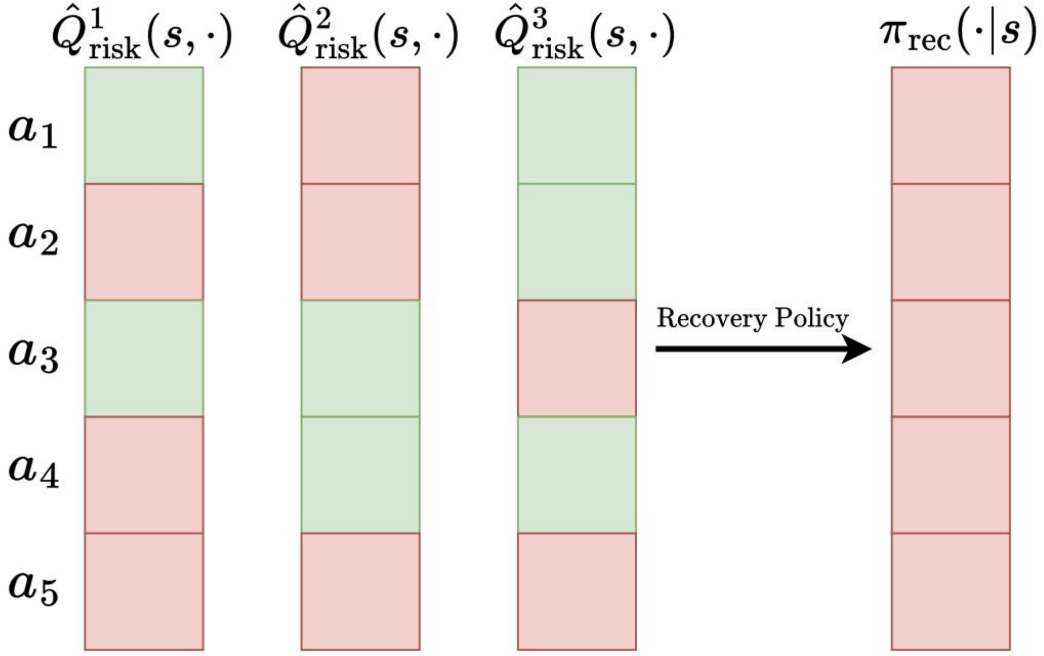


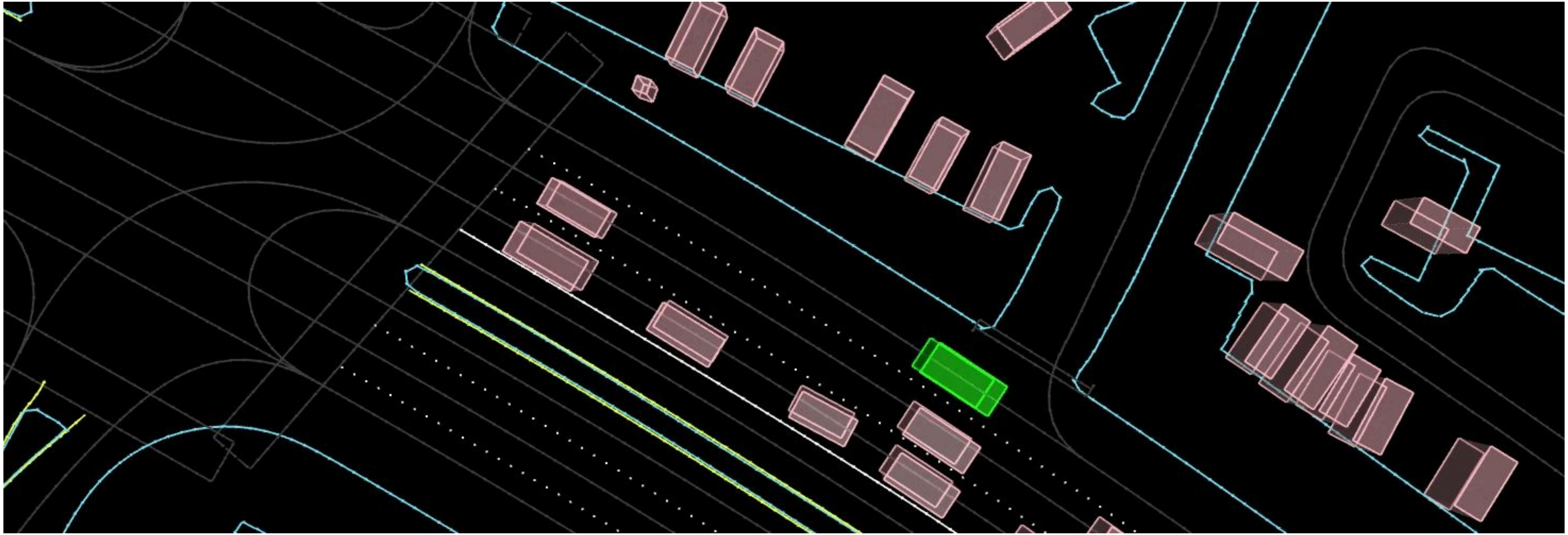
Constructing CIMRL Mixed Policy: Unsafe Case



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/

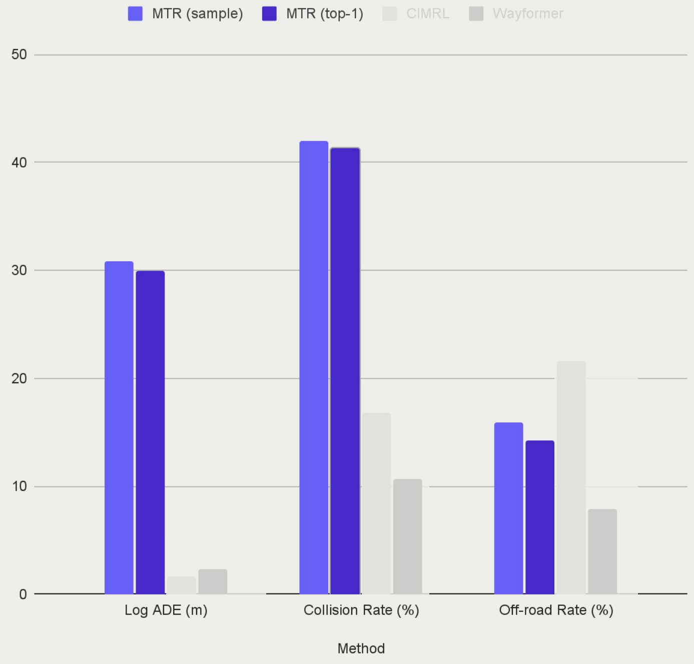
Gulino, Cole, et al. "Waymax: An accelerated, data-driven simulator for large-scale autonomous driving research." 2023.

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

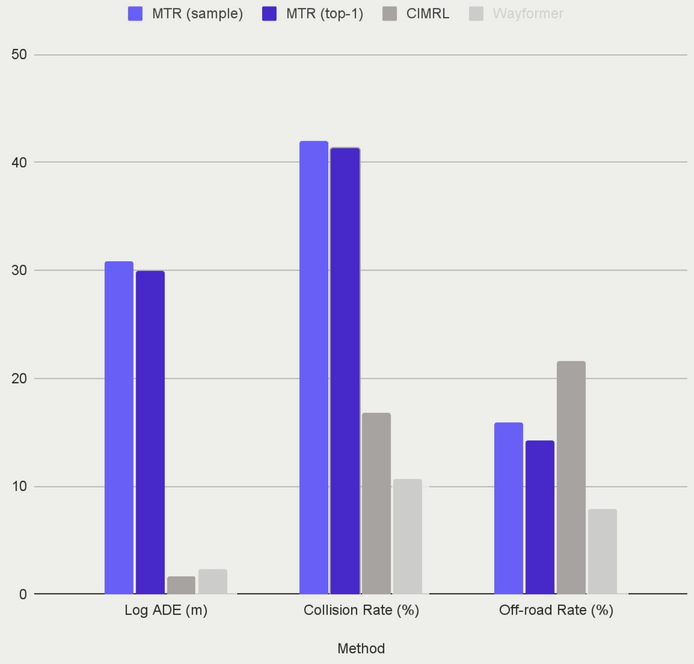


Closed-Loop Results: Waymax

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

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

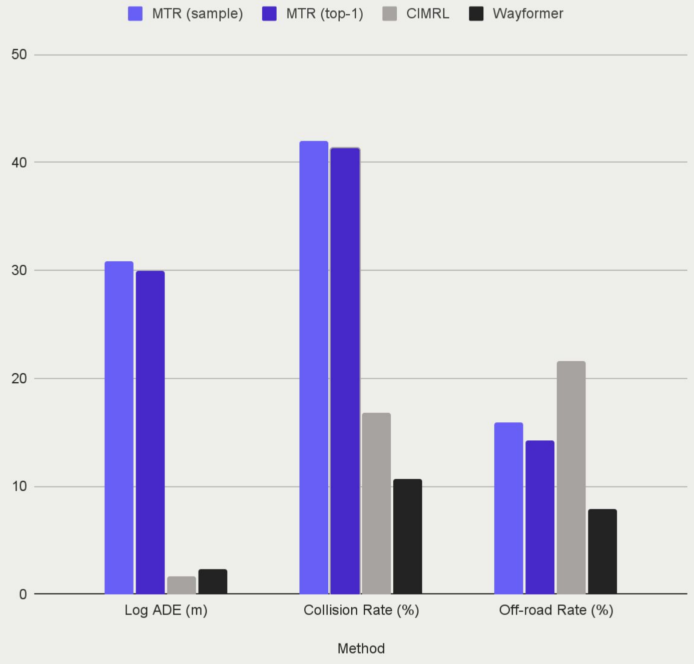
Using Waymax: No Sim Agents, Delta Action Space



Closed-Loop Results: Waymax

Wayformer has the access to route info :)

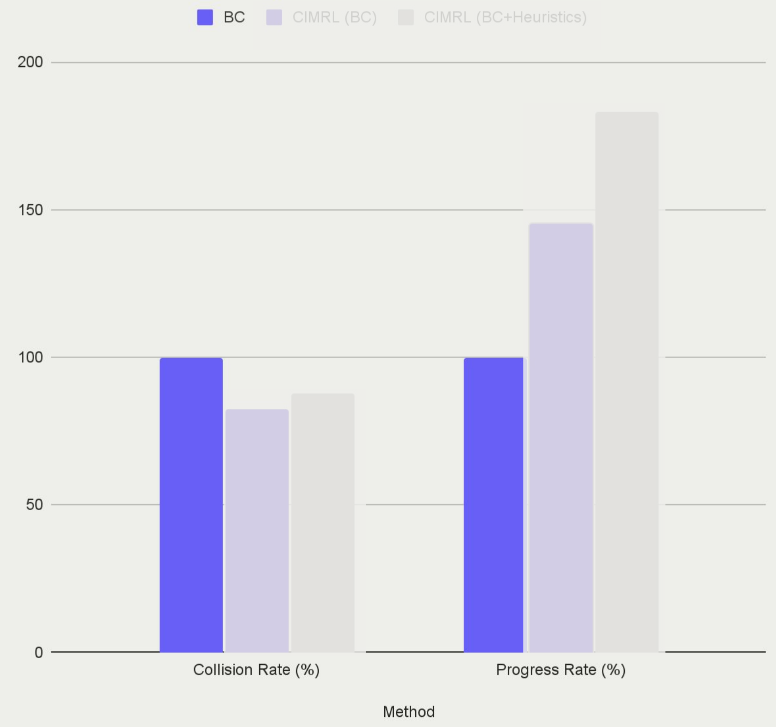
Using Waymax: No Sim Agents, Delta Action Space



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

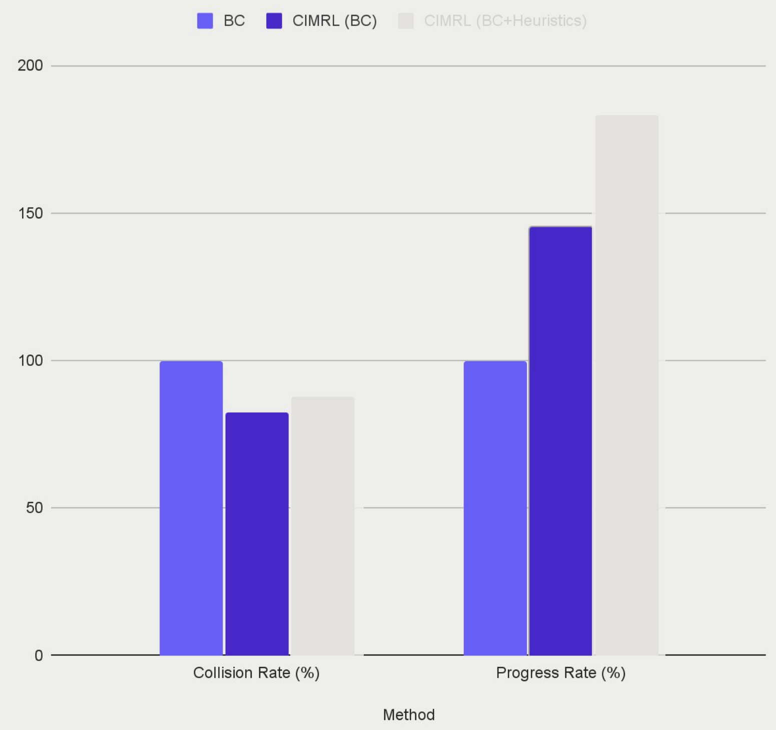
Using Internal data and Sim (Log replay)



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

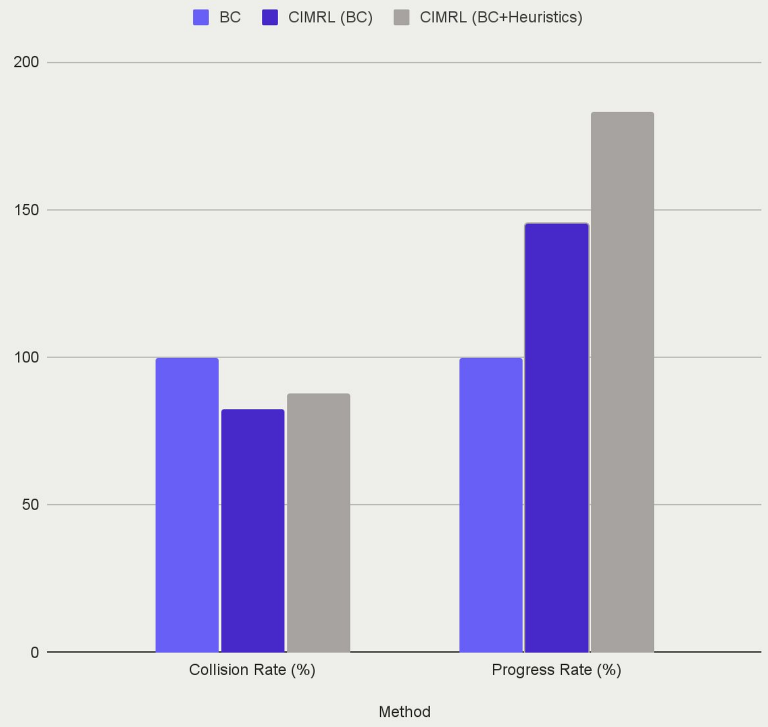
Using Internal data and Sim (Log replay)



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)



Conclusions

01

Logic-based reasoning helps with corner cases extractable from the rule-based KB

02

Learning selection provides long-horizon reasoning

03

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

Thanks!