iPrism: Characterize and Mitigate Risk by Quantifying Change in Escape Routes

Shengkun Cui, Saurabh Jha, Ziheng Chen, Zbigniew T. Kalbarczyk, Ravishankar K. Iyer

Is Autonomous Driving Safe Enough? [2018 – 2023]

03/2018

Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam

02/2020

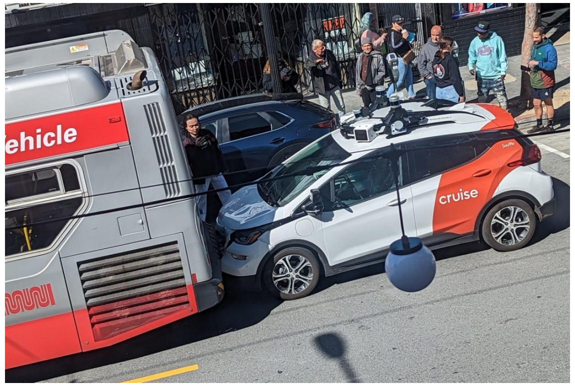
Apple Engineer Killed in Tesla Crash Had Previously Complained About Autopilot

By Tom Krisher and Olga Rodriguez
The Associated Press
Feb 11, 2020 Save Article

10/2023

Cruise Stops All Driverless Taxi Operations in the United States

The move comes just two days after California regulators told the company to take its autonomously driven cars off the road.



Is Autonomous Driving Safe Enough? [DSN 2018]

SAMPLE OF DISENGAGEMENT REPORTS FROM THE CA DMV DATASET.

Manufacturer	Raw Disengagement Report (Log)	Category	Tags
Nissan	1/4/16 — 1:25 PM — **Software module froze**. As a result driver safely disengaged and resumed manual control. — City and highway — Sunny/Dry	System	Software
Nissan	5/25/16 — 11:20 AM — Leaf #1 (Alfa) — The AV **didn't see** the lead vehicle, driver safely disengaged and resumed manual control.	ML/Design	Recognition System
Waymo	May-16 — Highway — Safe Operation — Disengage for a **recklessly behaving** road user	ML/Design	Environment
Volkswagen	11/12/14 — 18:24:03 — Takeover-Request — **watchdog error**	System	Computer System

We use the "—" to denote field separators.
Note that log formats vary across manufacturers and time.
Bold-face text represents phrases analyzed by the NLP engine to categorize log lines.

- AVs 15-4000x worse than humans
- Failures attributed to hardware/software, **uncertain environment** and ML for Waymo

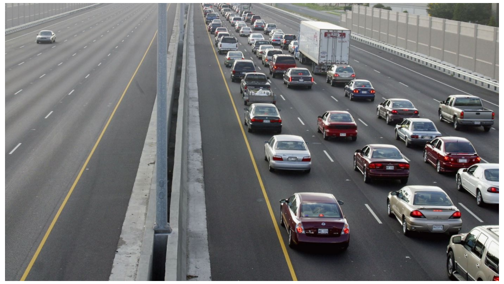
WAYMO

DSN 2018

How Do We Make Autonomous Driving Safer?

Why rear car choose to brake?

How Do We Make Autonomous Driving Safer?

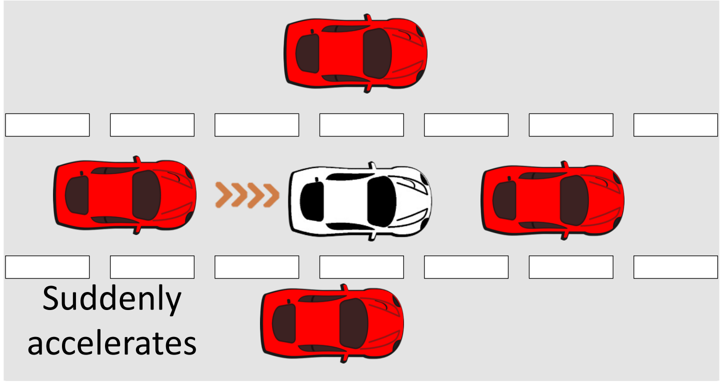


Attention required increases with the increase in uncertainty of another actor’s behavior

Ensuring Safety – Traditional Methods

- By avoiding collision trajectories
  - Time to collision
  - Intel Responsibility Sensitive Safety (RSS)
  - Nvidia Safety Force Field (SFF)

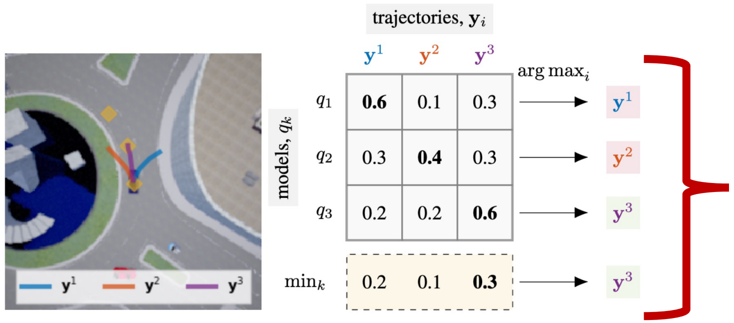
- **Does not proactively assess risks**
  - **Predicted collision trajectories can be inaccurate**
  - **Often too late to avoid accident**



RSS/SFF cannot avoid accident!

- By learning from data
  - Reinforcement learning
  - Imitation learning
  - Adaptation to out-of-training-distribution

- **Depends on training data quality**
  - **Data inefficiency: require large amount of training data**
  - **Cannot handle rare driving scenarios**



(a) OOD driving scenario (b) Robust imitative planning

What if y^1, y^2, y^3 are all unsafe?

RIP agent (Filos et. al.) crashes under an OOD scenario in CARLA simulation

How to overcome these shortcomings?

Handle Uncertainty via Safe Back-up Plans

- Uncertainties always exist in practice!
  - Sensor/SW/HW faults and failures
  - Less robust ML model prediction in out-of-training-distribution scenarios
  - **Unpredictable Behavior of other actors**

- What can we do then?
  - Ensuring enough back-plans (aka escape routes)
  - **Maximizing the chance of having safe routing choices (in uncertain environment)**

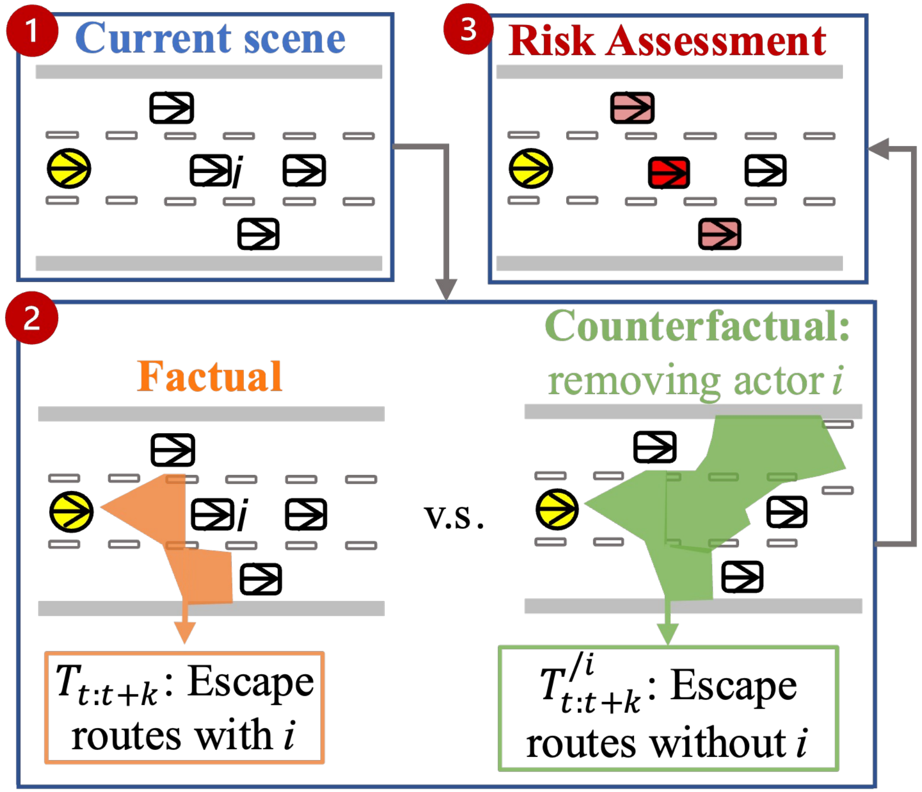
AD Safety & Risk Assessment

Human intuitions

1. Actively ensure "backup plans" (aka "escape routes")
2. Handle uncertainty and zero-day scenarios

Research Question 1:
How do we design a risk metric that embeds these intuitions?

Analytical, no learning needed!



$$STI_i \propto |T^{/i}_{t:t+k}| - |T_{t:t+k}|$$

Motivated from Barlow & Proschan work [1975]

Risk Assessment in Practice

1. Compute escape routes via reachability analysis

$$\underbrace{\mathcal{T}_{t:t+k}}_{\text{Escape Routes}} = f\left(\underbrace{x_t^{(ego)}, \mathcal{X}_{t:t+k}}_{\text{Actor states and trajectories}}, \underbrace{\mathcal{M}}_{\text{Drivable Area}}, \underbrace{\mathcal{D}, u}_{\text{Vehicle dynamic model and possible control inputs}}\right)$$

$f(\cdot)$: a reachability analysis algorithm

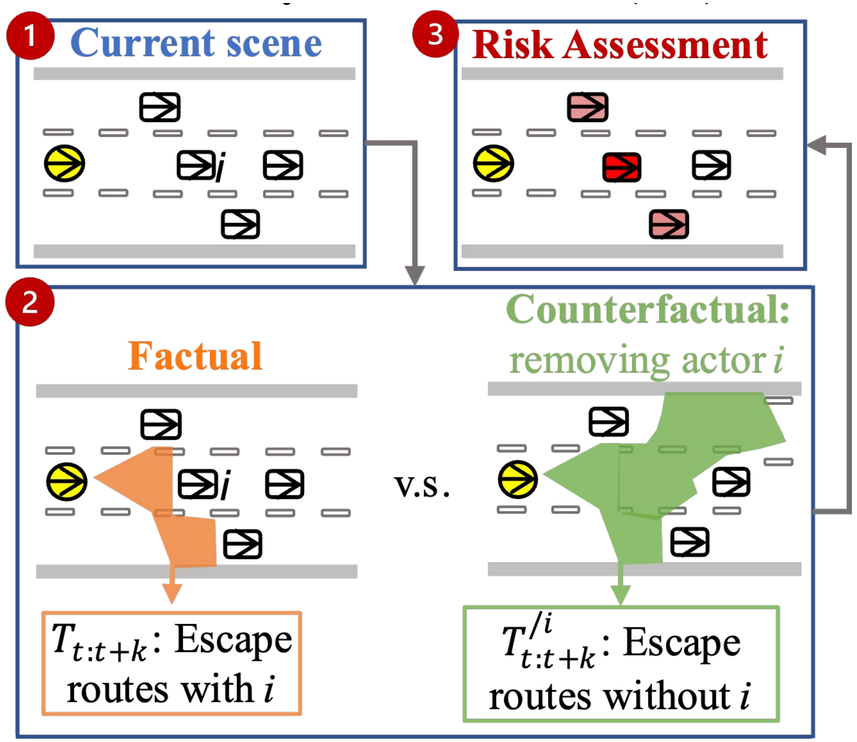
2. Compute reach-tube with actor removal (counterfactual)

$$\mathcal{T}^{\boxed{/i}}_{t:t+k} = f\left(x_t^{(ego)}, \mathcal{X}^{/i}_{t:t+k}, \mathcal{M}, \mathcal{D}, u\right)$$

$$\mathcal{T}^{\boxed{\varnothing}}_{t:t+k} = f\left(x_t^{(ego)}, \mathcal{X}_{t:t+k} = \varnothing, \mathcal{M}, \mathcal{D}, u\right)$$

3. Compute STI (risk) value

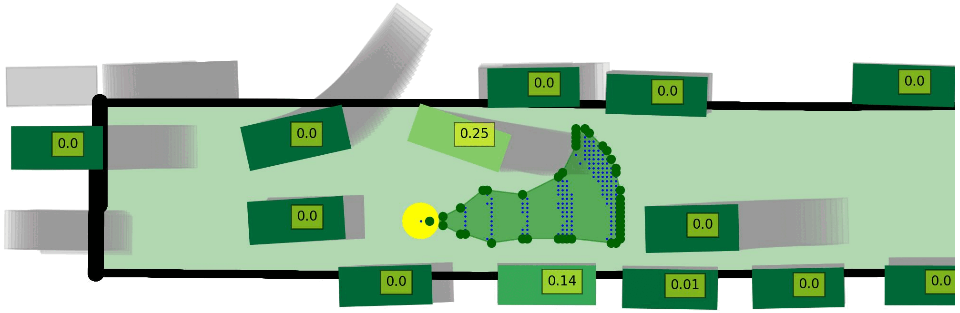
$$STI_t^{(i)} = \frac{|\mathcal{T}^{/i}_{t:t+k}| - |\mathcal{T}_{t:t+k}|}{|\mathcal{T}^{\varnothing}_{t:t+k}|} \qquad STI_t^{(combined)} = \frac{|\mathcal{T}^{\varnothing}_{t:t+k}| - |\mathcal{T}_{t:t+k}|}{|\mathcal{T}^{\varnothing}_{t:t+k}|}$$

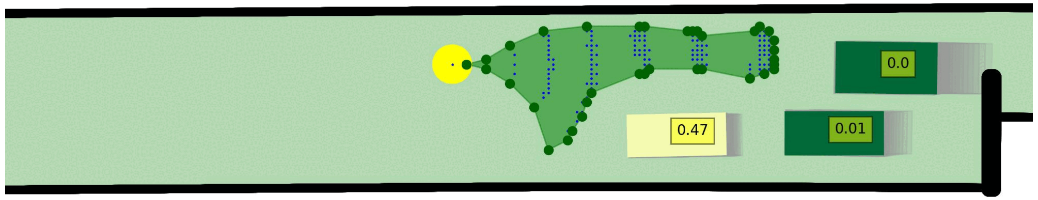


$$STI_i \propto |T^{/i}_{t:t+k}| - |T_{t:t+k}|$$

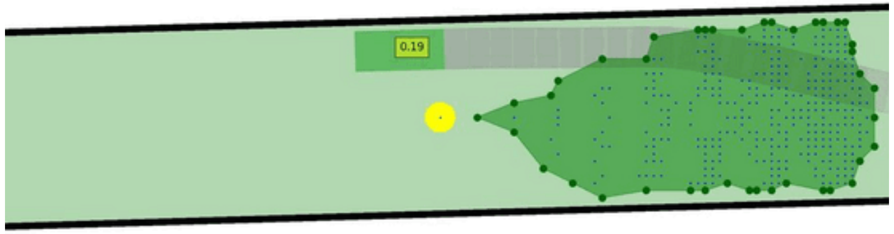
Demonstration of Risk Assessment

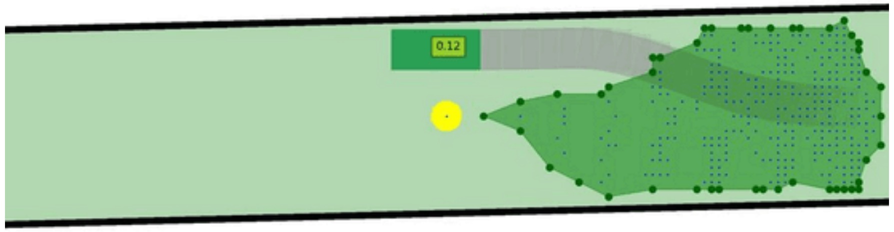
Argoverse (Chang et al. 2019) Real-world Dataset



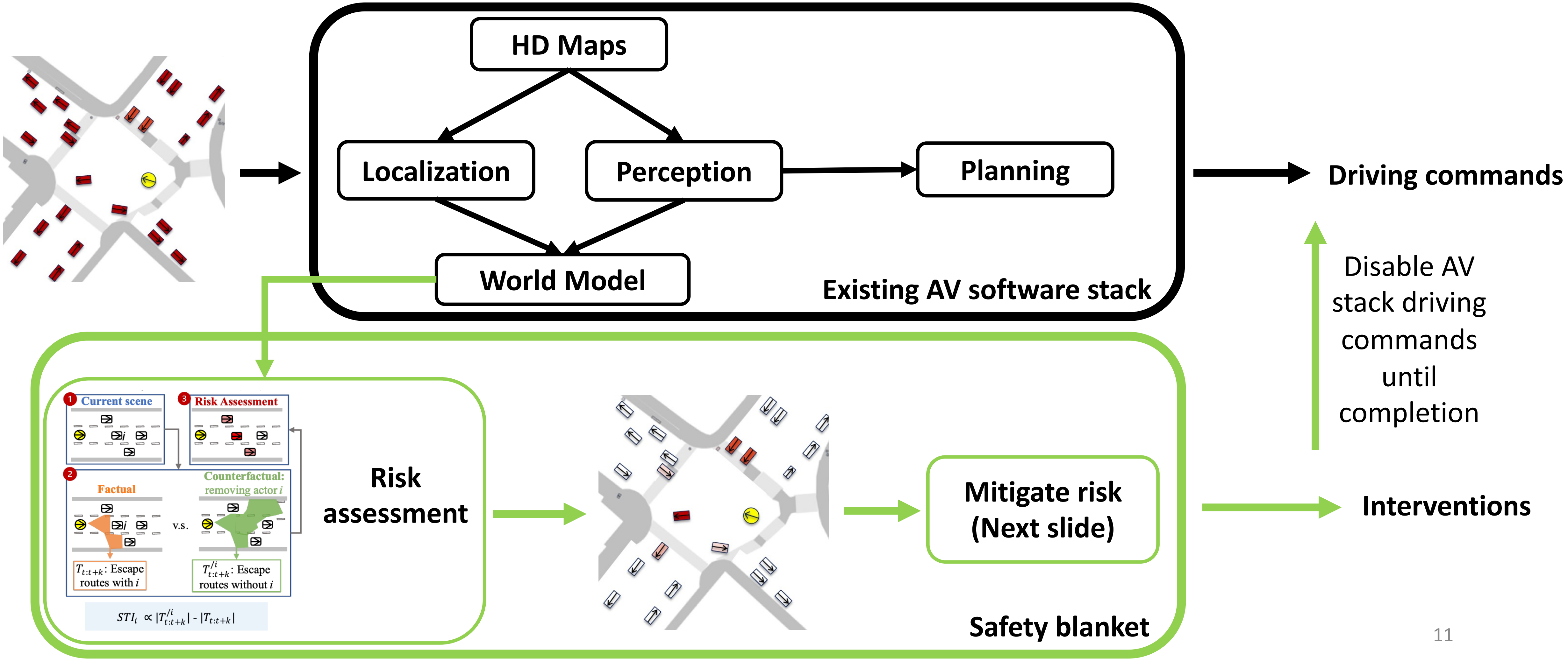


CARLA Simulator with High-risk OOD Scenarios





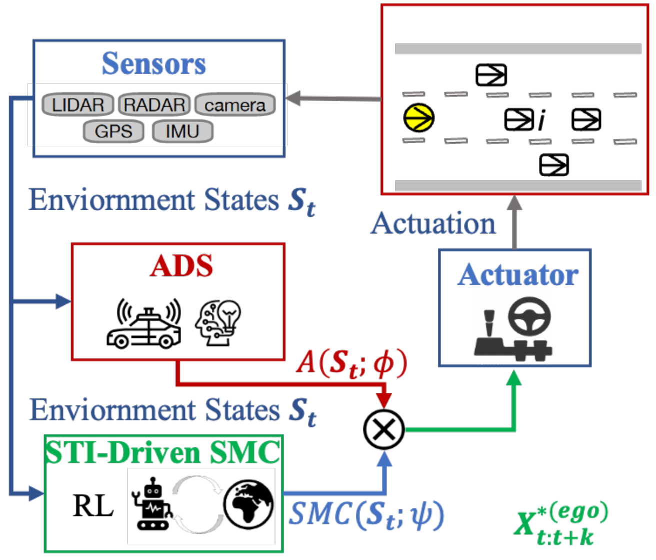
Risk Metric Application: Risk-aware Safety Blanket



Risk-driven Mitigation with RL

Risk (STI) reduction via mitigation

1. Safety-hazard mitigation controller (SMC) acts (policy) to reduce the STI
2. Learn mitigation policy via RL
3. STI is part of the reward during training



$$X^{*(ego)}_{t:t+k} = A(S_t;\phi) \otimes SMC(S_t;\psi)$$

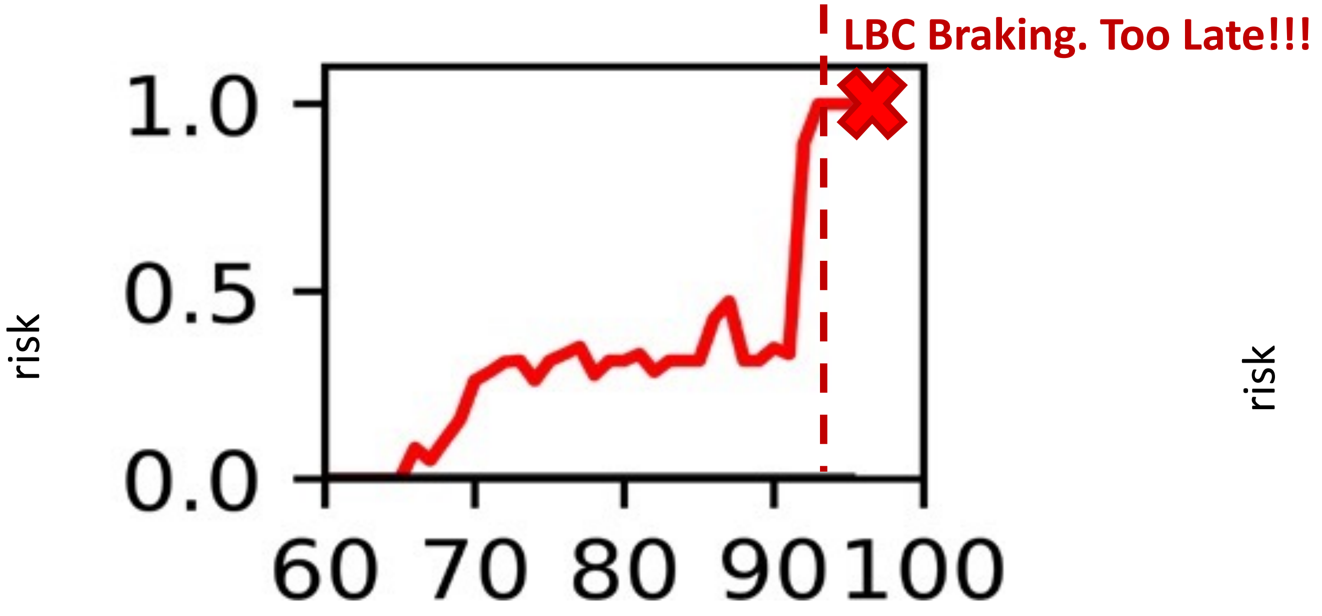
Research Question 2:
How do we use the risk metric to provide mitigation actions?

S_t: Sensor data (e.g., camera frames)

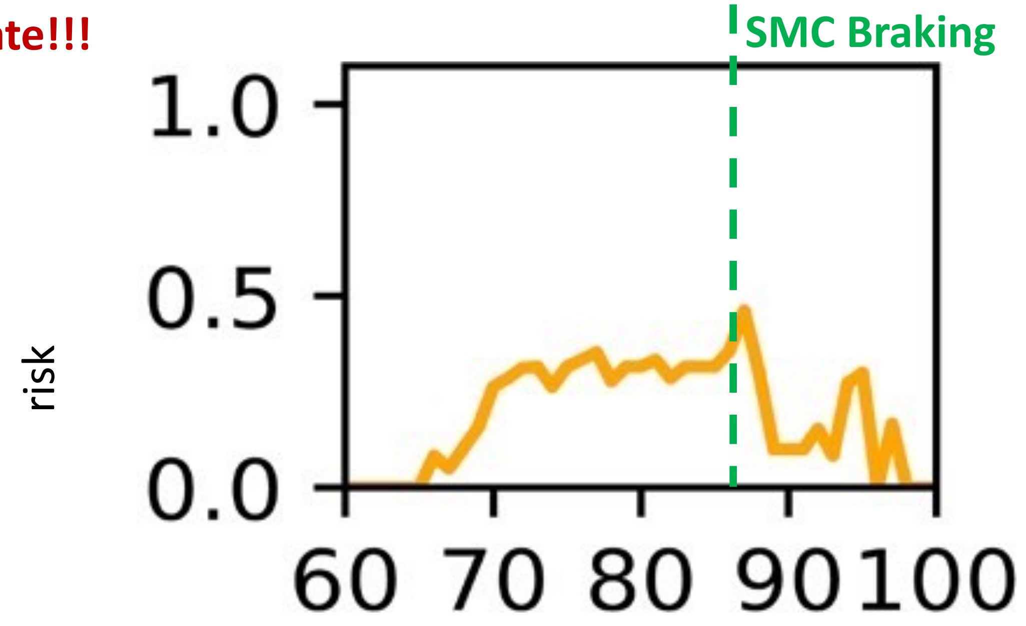
a_t: Mitigation action (e.g., braking, changing lane)

R: STI-driven reward model
(e.g., $r_t = \alpha_0(1 - STI) + \alpha_1 \texttt{GoalCompletionTerms}$)

Proactive Reduce Risk for Mitigating Accidents



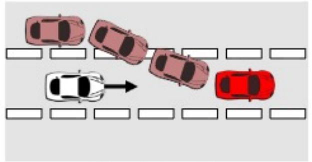
Baseline (LBC agent by Chen et al. 2020)

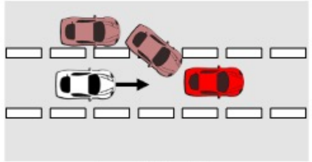


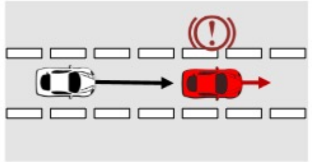
Ours (LBC agent + STI-based SMC)

Proactively avoids **trajectories of no return** by reducing risk!

Results







Agent	Ghost cut-in	Lead cut-in	Lead slowdown
LBC + Ours	**267**	**3**	**15**
LBC	519	170	118

Agent	Ghost cut-in	Lead cut-in	Lead slowdown
RIP + Ours	**65**	**265**	**129**
RIP	478	671	440

 # collisions in 1000 scenarios per typology (lower is better)

Significant reduction in accidents

Conclusion and Future Work

- **Conclusion**
  - Defining risk metric that captures escape routes and use it for remediation

- **Future work**
  - How to apply such techniques in cloud resilience?
    - Risk assessment, Root cause analysis, Remediation

  - How can modern BN + LLMs (trained on TBs of data) help?
    - Identify key system events in risk state from system logs and metric data?
    - Auto-correlates failure events that ultimately lead to SWO?
    - Remediation action recommendation and activation?