Safety Assurance of Autonomous Systems

Shreyas Ramakrishna (shreyas.ramakrishna@vanderbilt.edu)

Graduate Research Assistant @ scopelab

Institute For Software Integrated Systems, Vanderbilt University

https://www.shreyasramakrishna.com/

About Me

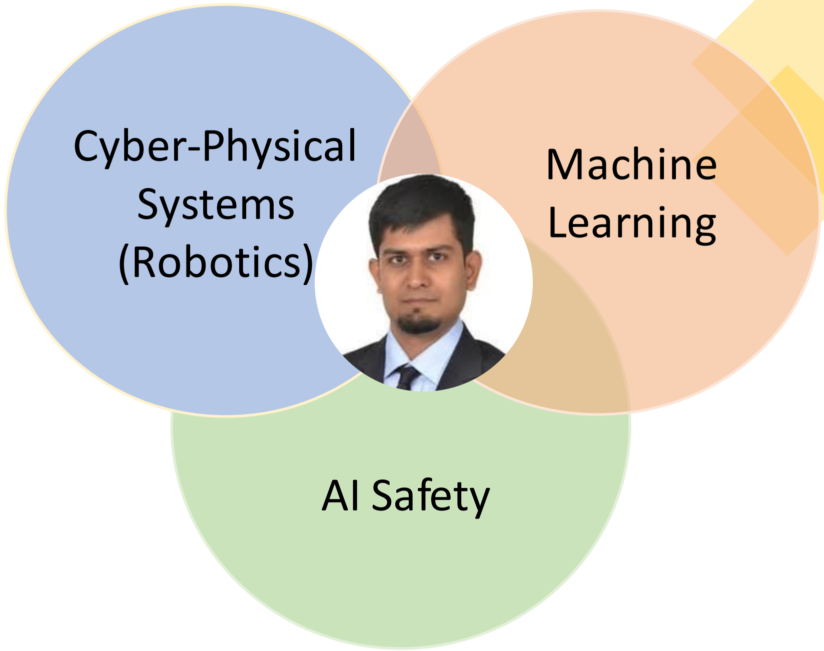
Education

- Currently, A fifth year Ph.D. candidate @ Institute For Software Integrated Systems, Vanderbilt University, working @ scope lab with Professor Abhishek Dubey for DARPA's Assured Autonomy Project. (2017-Present)

- Masters in Electrical Engineering from Technical University Kaiserslautern (Germany), with Master Thesis @ Department of Cyber-Physical Systems. (2013-2015)

Work Experience

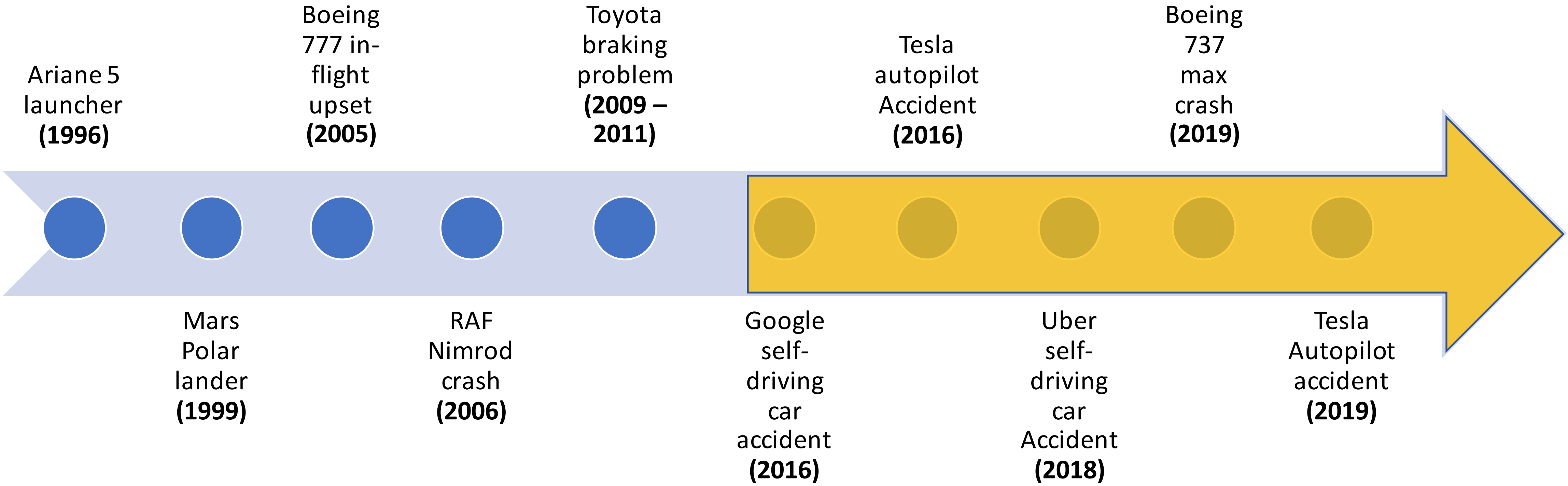
- Summer Research Intern @ Siemens Corporate Technology, Princeton, NJ. (May – Aug 2021)
- Embedded Design Engineer @ Apsis Solutions, Bangalore, India. (2015-2017)



Autonomous Cyber-Physical Systems

AI and ML have revolutionized the field of Cyber-Physical Systems. But assuring the safety of these systems is a challenge.

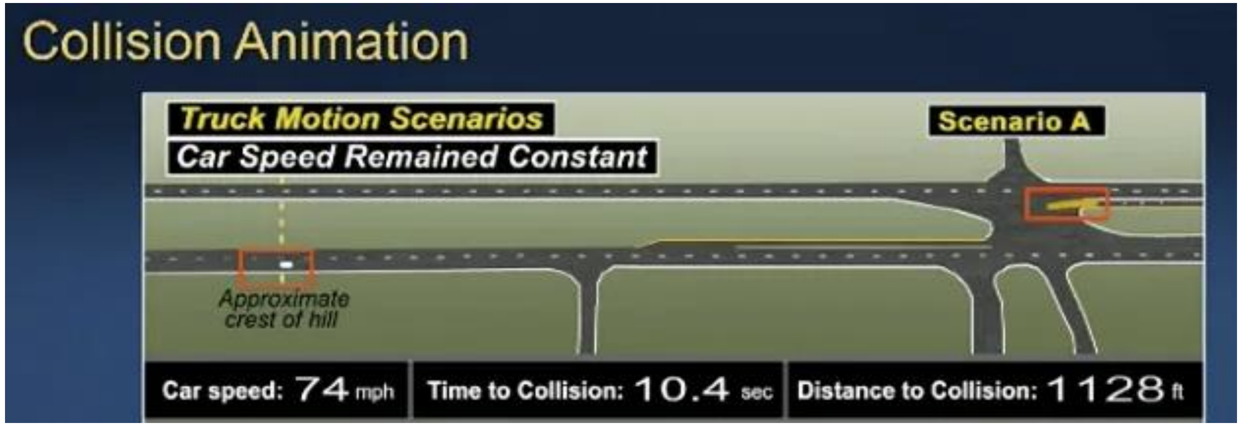
Timeline of Several Critical Accidents



Accidents are increasing with the need to achieve higher levels of autonomy

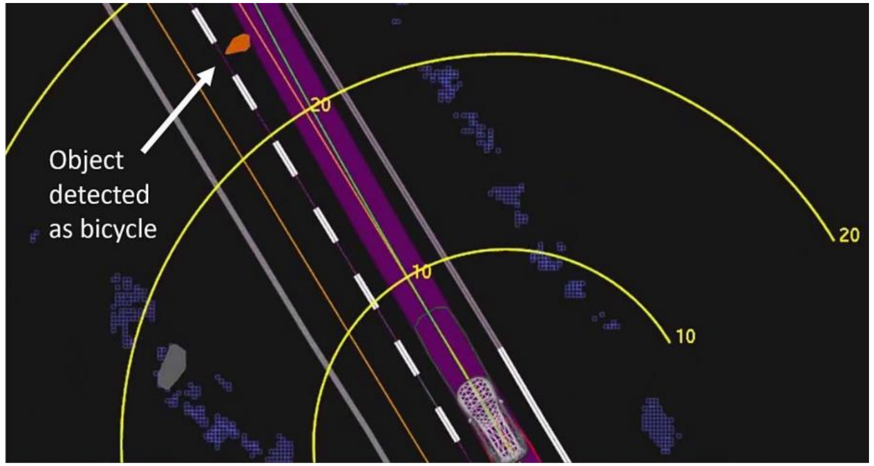
Accident Reports of Autonomous Vehicles

Tesla Accident – May 2016



"The Tesla's automated vehicle control system was not designed to, and did not, identify the truck crossing the car's path or recognize the impending crash"
- NTSB Report 2017

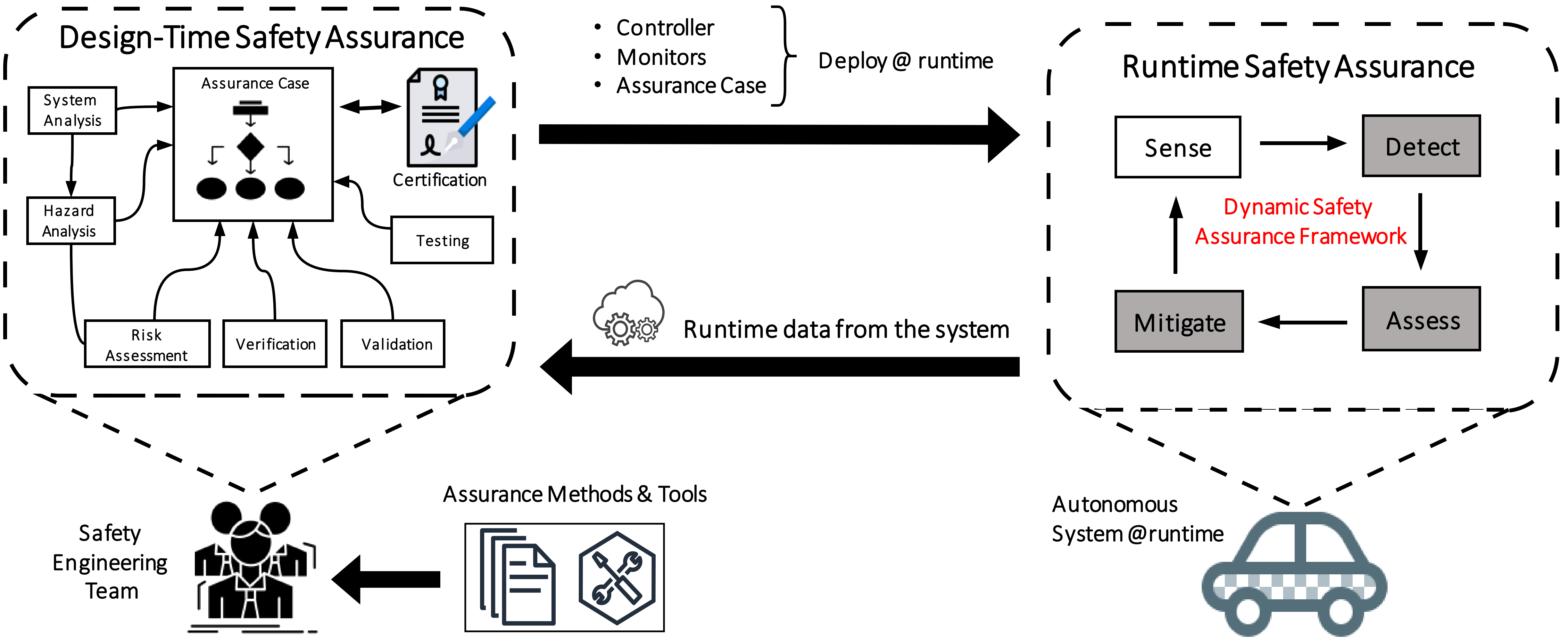
Uber Accident – March 2018



"The self-driving system software classified the pedestrian as an unknown object, as a vehicle, and then as a bicycle with varying expectations of future travel path. At 1.3 seconds before impact, the self-driving system determined that an emergency braking maneuver was needed"
- NTSB Report 2018

1. TESLA Crash report https://www.ntsb.gov/investigations/accidentreports/reports/har1702.pdf
2. Uber Accident report https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf

Research Overview – Dynamic Safety Assurance

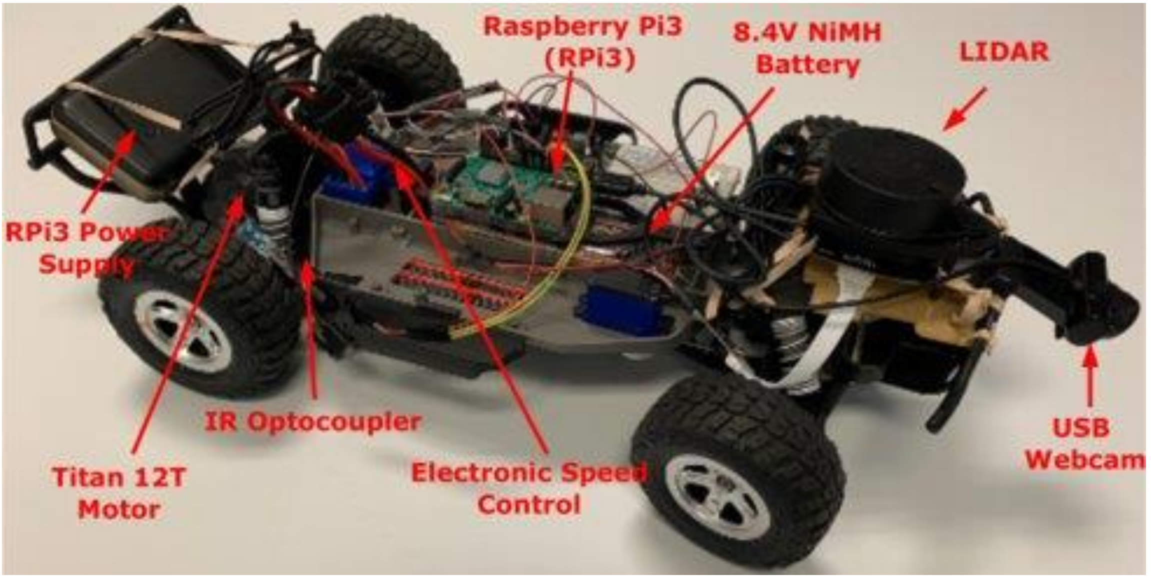


Current approaches primarily focus on safety @design-time

My research focus on safety @runtime

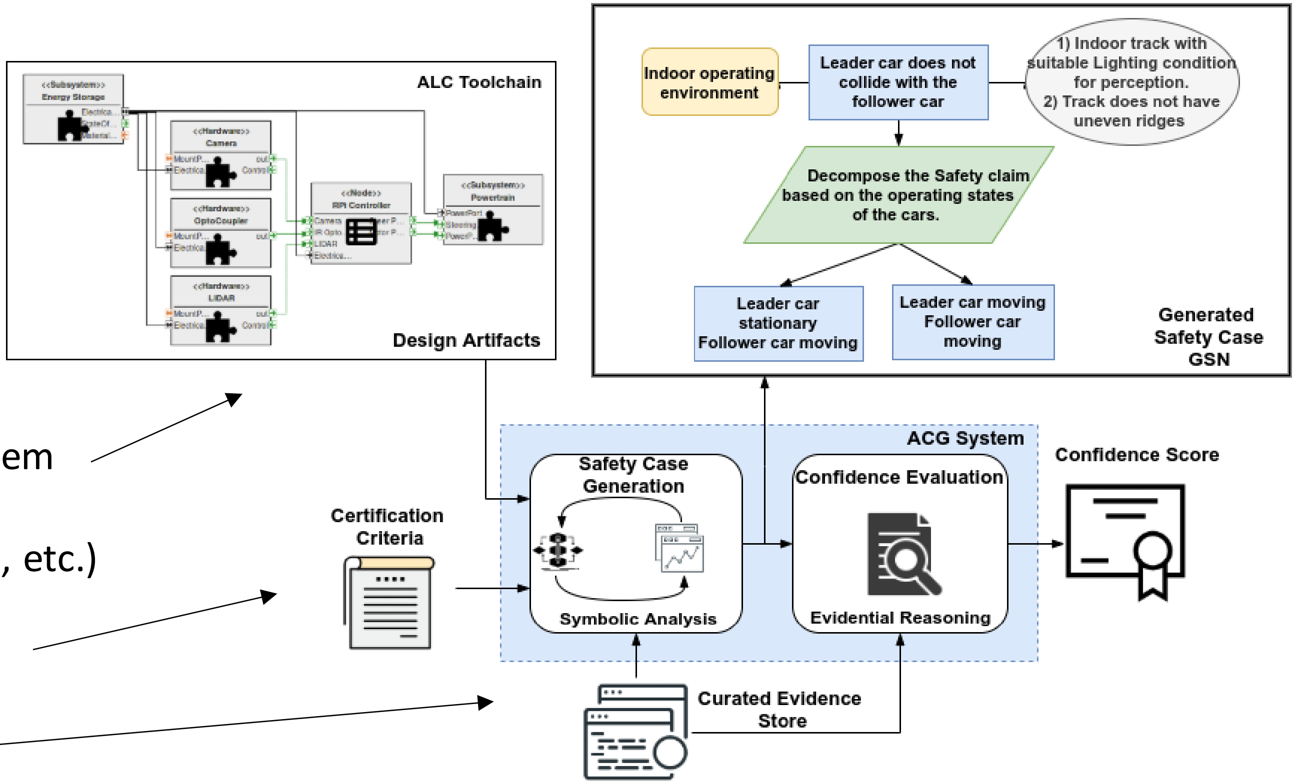
Demonstration Platforms to Validate My Research

Carla Simulation

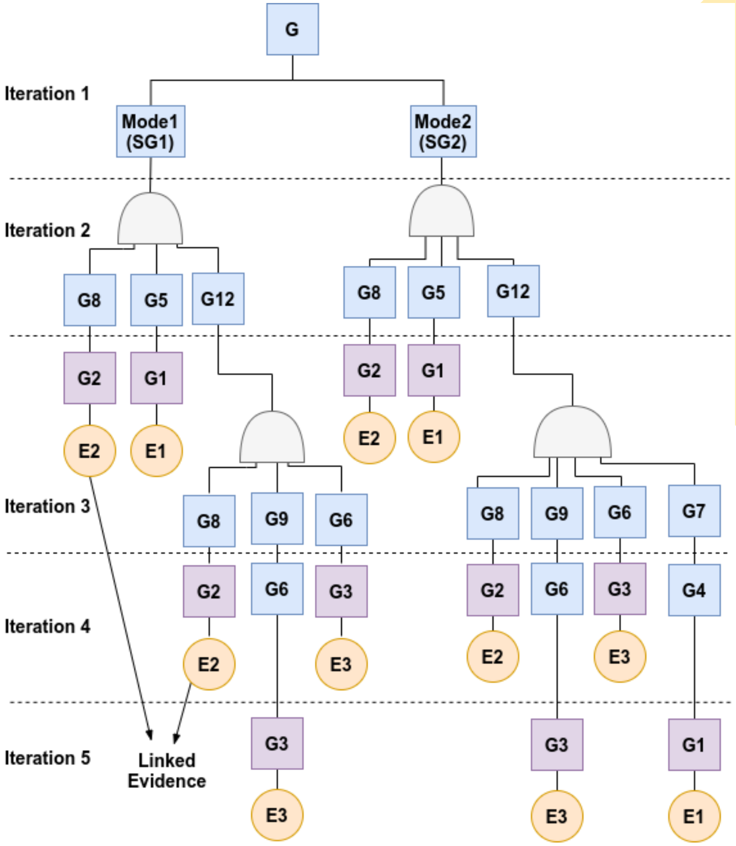


DeepNNCar Autonomous Driving Testbed

Design-Time Assurance Case (DARPA ARCOS)



1. Design Artifacts (system architecture models, interconnectivity graph, etc.)
2. Certification criteria
3. Evidence store



Output Goal Structuring Notation

Assurance Case is a structured argument set backed by system evidence to prove that the system will safely operate in a given environment.

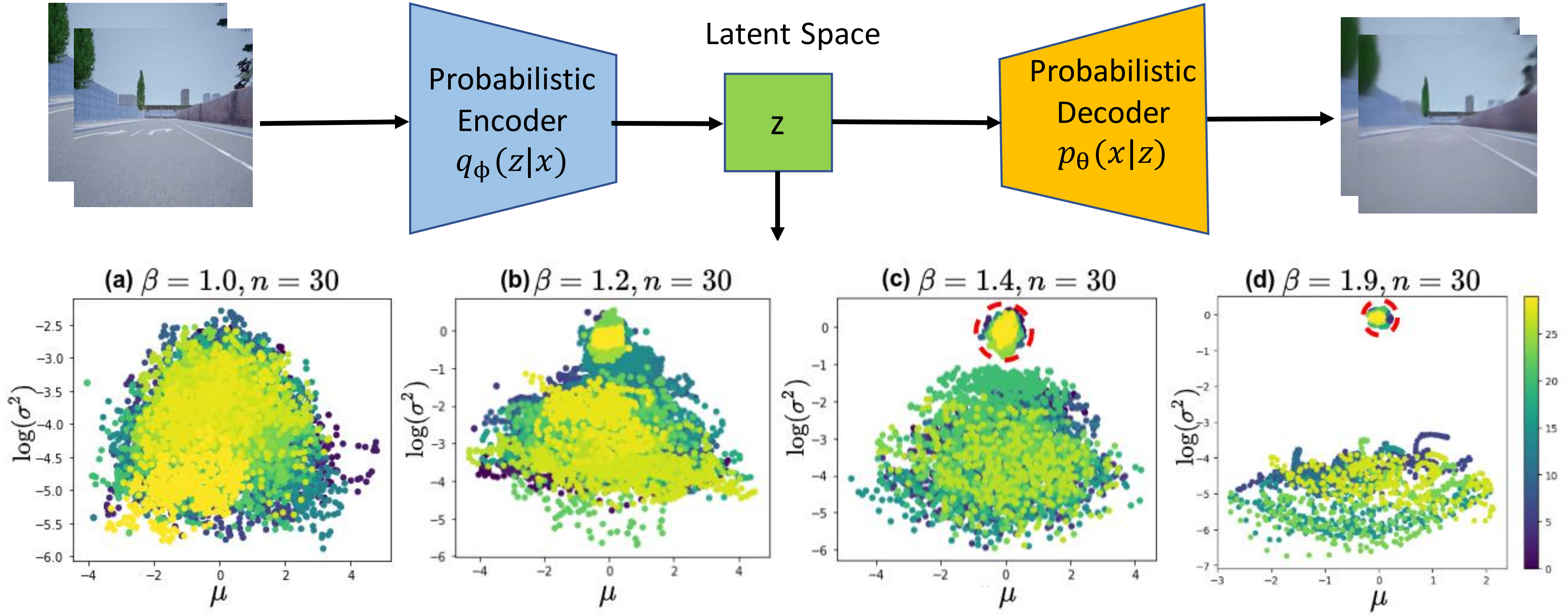
Out-of-Distribution Problem

Shift in Operating Conditions

Adversarial Attacks

Machine Learning components are trained under a closed world assumption that the operational data is drawn IID with the training dataset, which is not always true. – Out-of-Distribution Problem

Unsupervised Out-of-Distribution Detection



Hyperparameters
N is the number of latent variables
β controls data flow

Key Concept - We systematically tune the hyperparameters of β-VAE network to **partially disentangle the latent space** and then **learn an approximate mapping** of the input to the latent variables to perform OOD detection.

Publications

1. **Ramakrishna, S.**, Rahiminasab, Z., Easwaran, A., & Dubey, A. (2020, September). "Efficient Multi-Class Out-of-Distribution Reasoning for Perception Based Networks: Work-in-Progress." In *2020 International Conference on Embedded Software (EMSOFT)*
2. **Ramakrishna, S.**, Rahiminasab, Z., Karsai, G., Easwaran, A., & Dubey, A. (2021). "Efficient Out-of-Distribution Detection Using Latent Space of β-VAE for Cyber-Physical Systems." in TCPS 2020

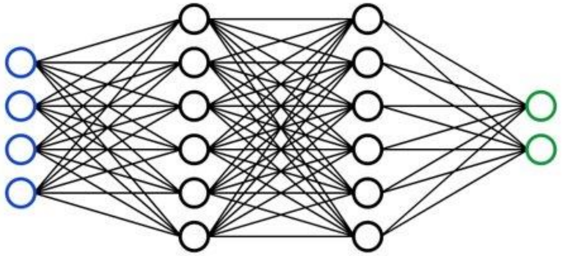
Runtime Verification Detector (Reachability Analysis)

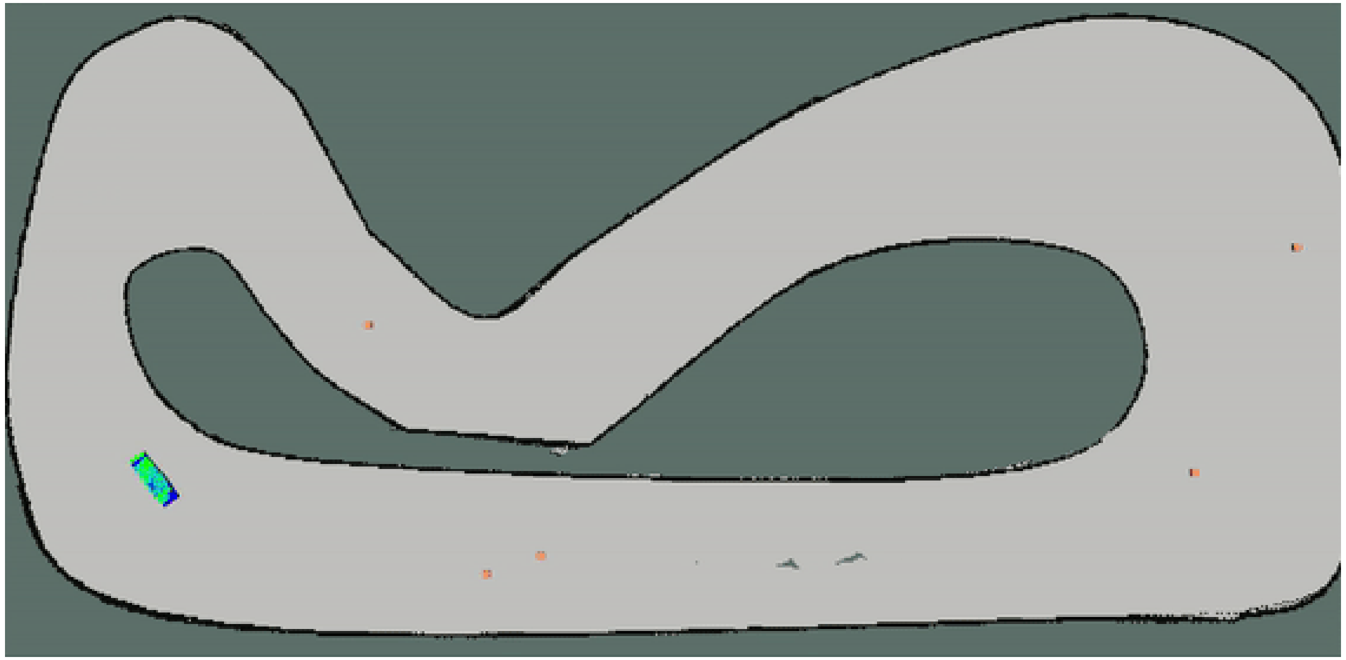
SysID Model

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t))$$

x, y, yaw speed

Controller, steering, throttle





- System properties to be proved is expressed as runtime monitors
- Safe states identified using reachability analysis

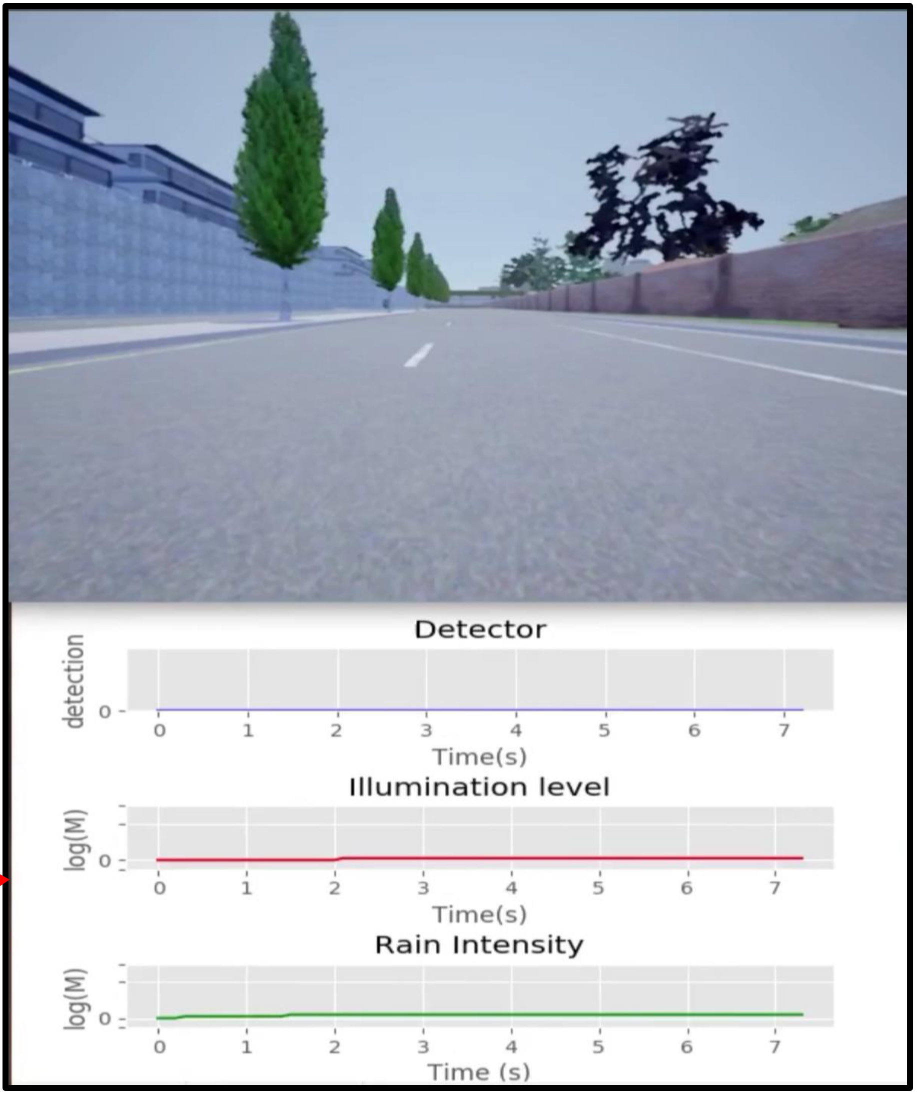
Out-of-Distribution Detection Demonstration in CARLA Simulation

Scenario1: Changing Rain intensity

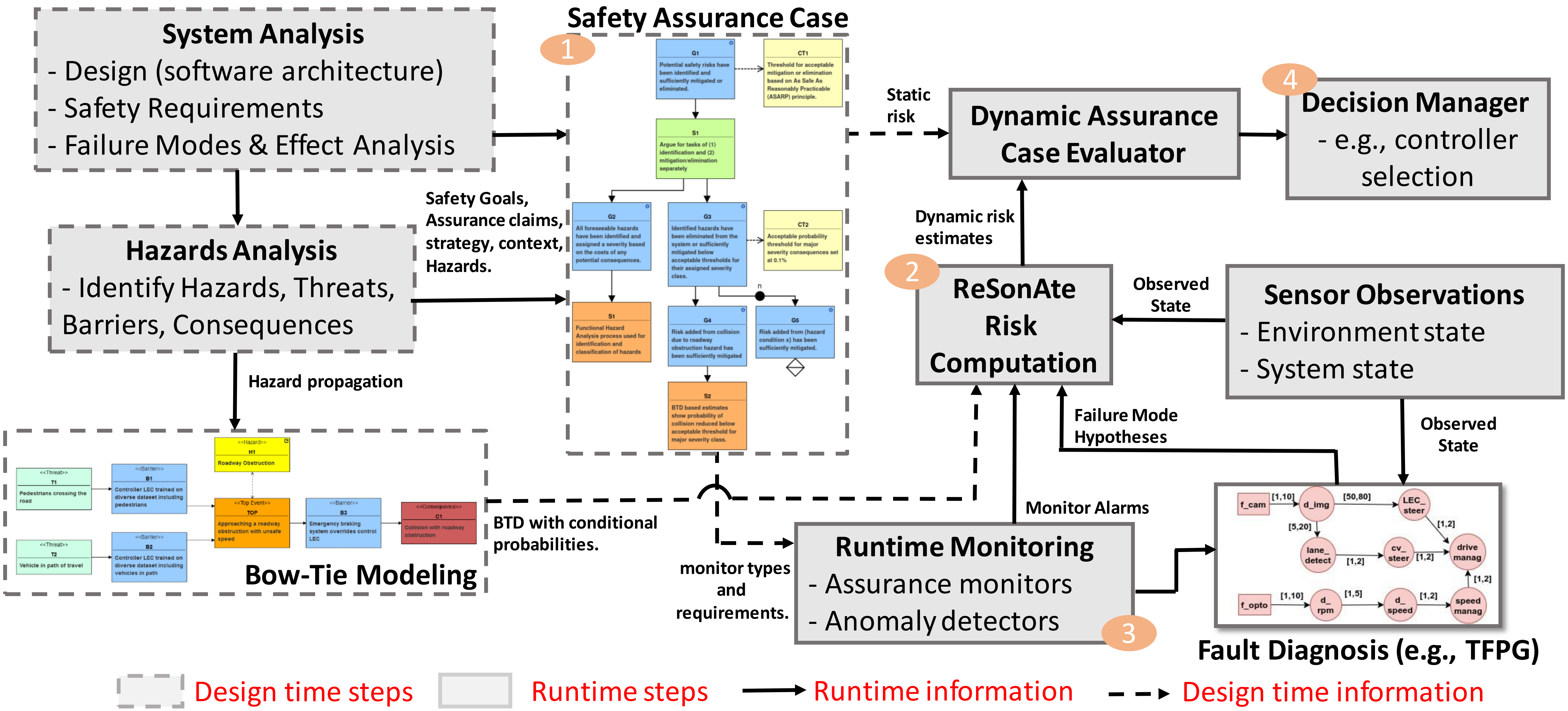
At t = 13s, the perception is increased to 60%. **(OOD scenario)**

Scenario2: Changing illumination level

At t = 13s, the light illumination is increased above 70 lumens. **(OOD scenario)**

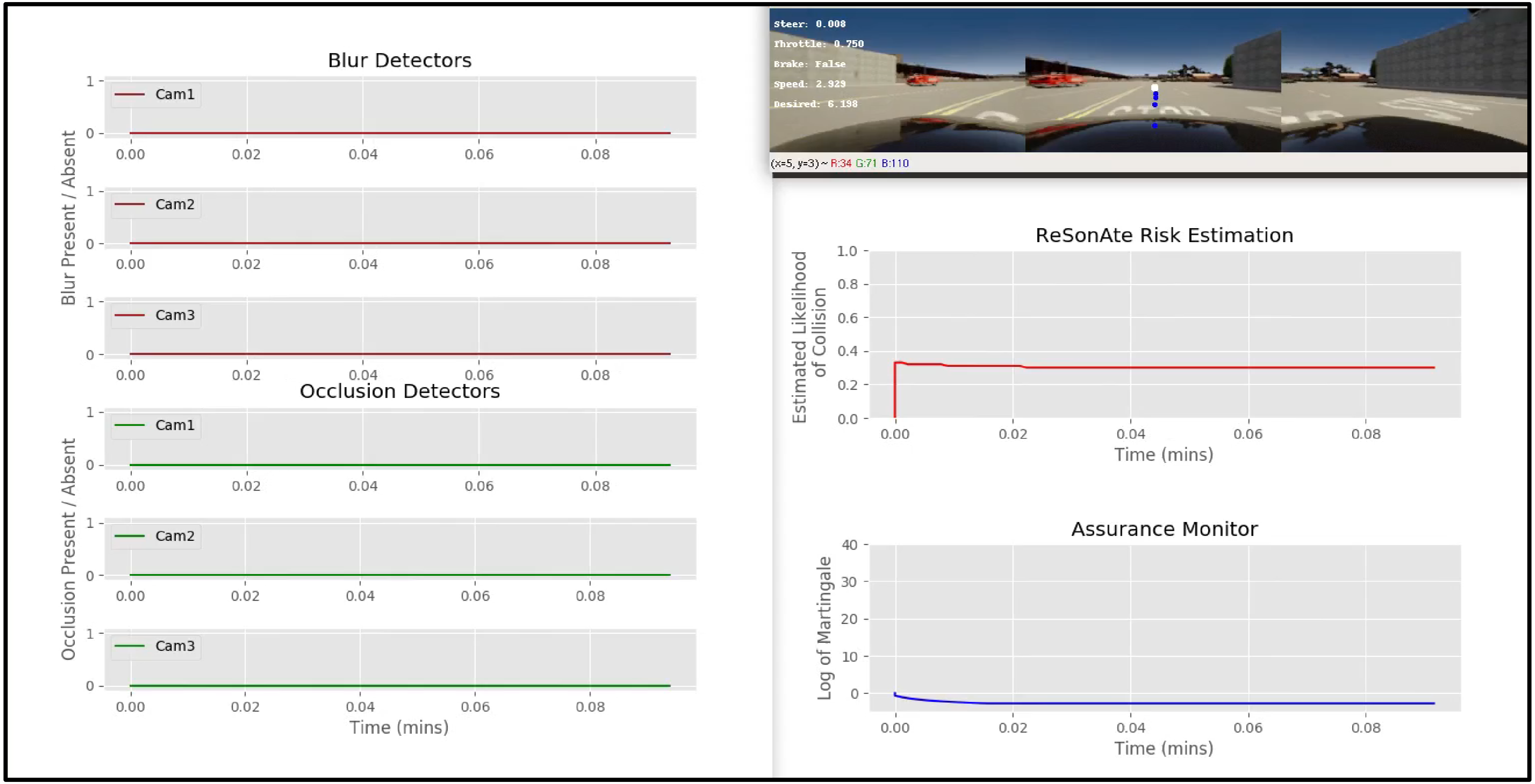


Risk Assessment Framework for Autonomous Systems

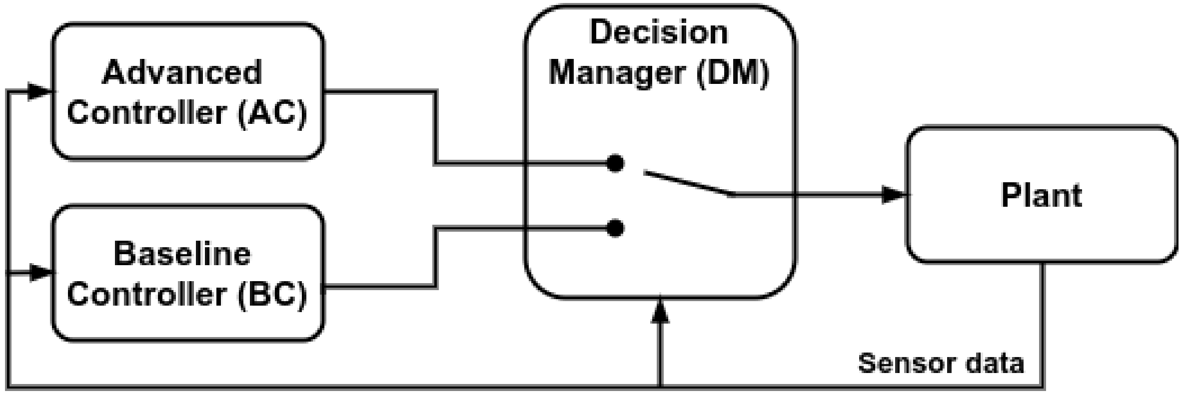


Github - https://github.com/scope-lab-vu/Resonate

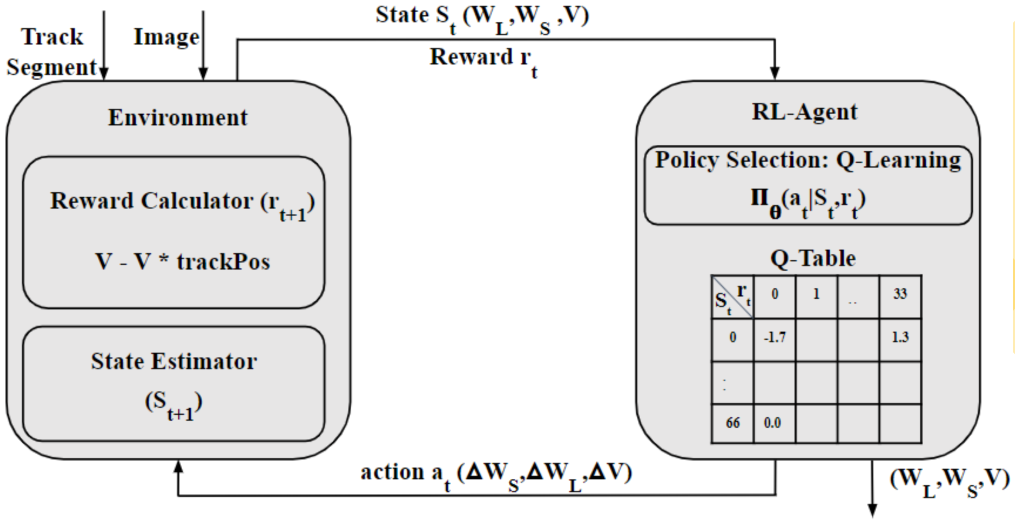
Risk of AV with Camera Failure



Runtime Decision Making under uncertainty



Conventional Simplex Architecture



Reinforcement Learning Setup

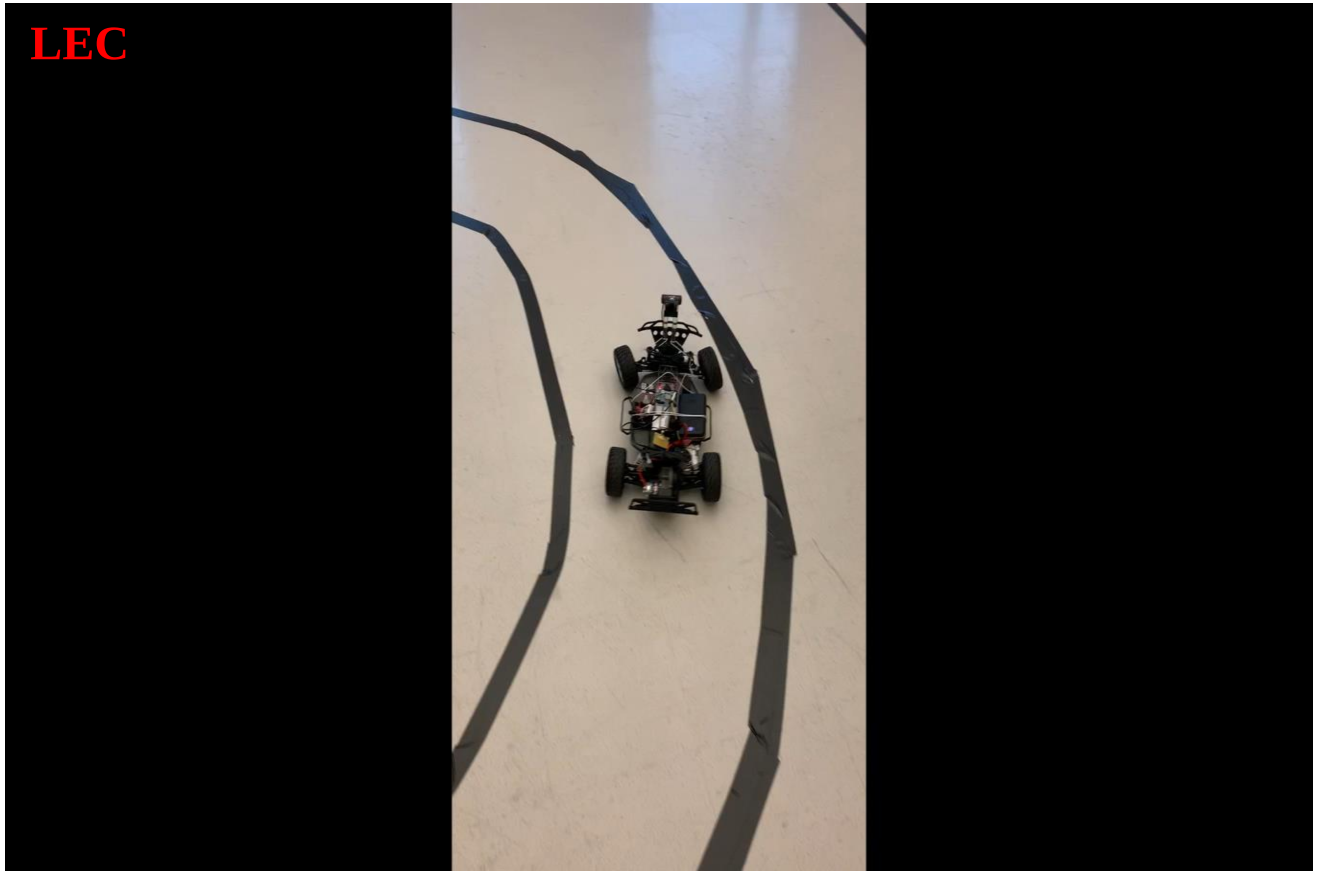
- **Problem** - Decision logic is learned offline. However, it needs to be "Proactive" and "Adaptive".
- Reinforcement learning (Q-learning) to learn dynamic decision weights.

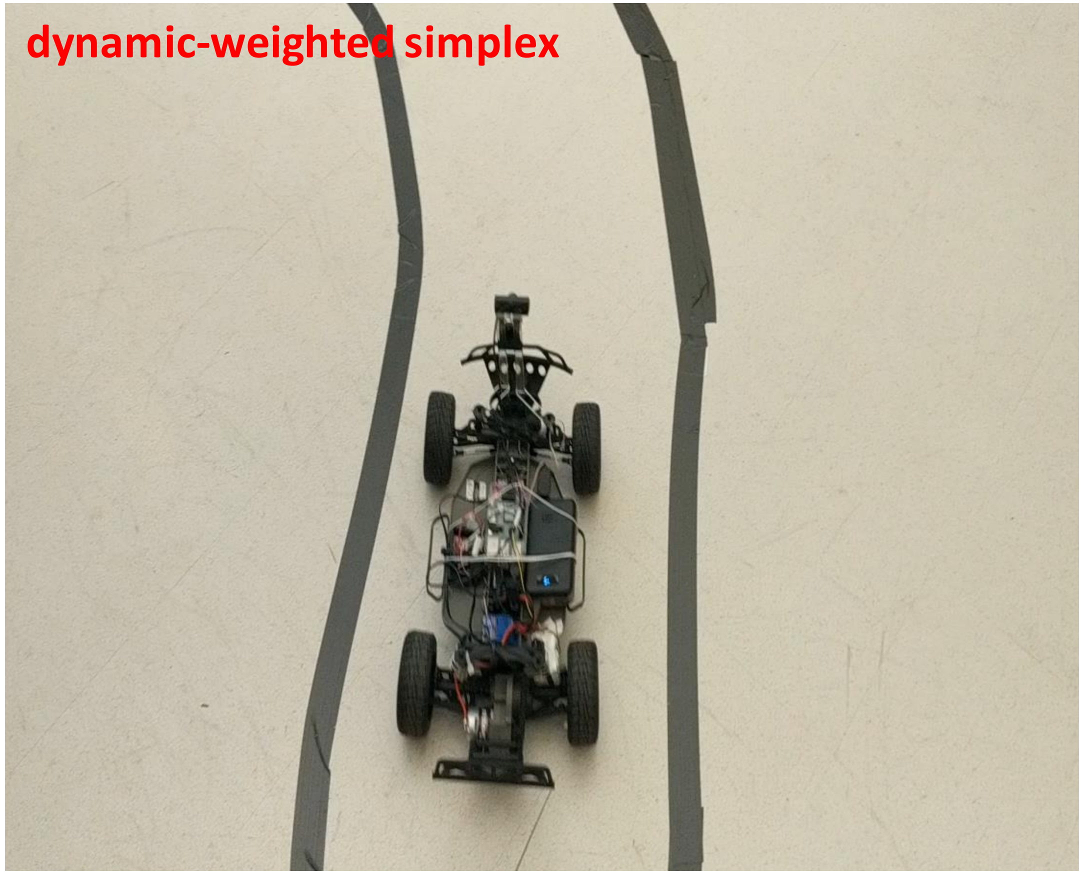
Publications

1. Ramakrishna, Shreyas, et al. "Dynamic-weighted simplex strategy for learning enabled cyber physical systems." *Journal of systems architecture* 111 (2020): 101760.

2. Ramakrishna, Shreyas, et al. "Augmenting learning components for safety in resource constrained autonomous robots." *2019 IEEE 22nd International Symposium on Real-Time Distributed Computing (ISORC)*. IEEE, 2019.

Crossing Track Boundaries – DeepNNCar Demo

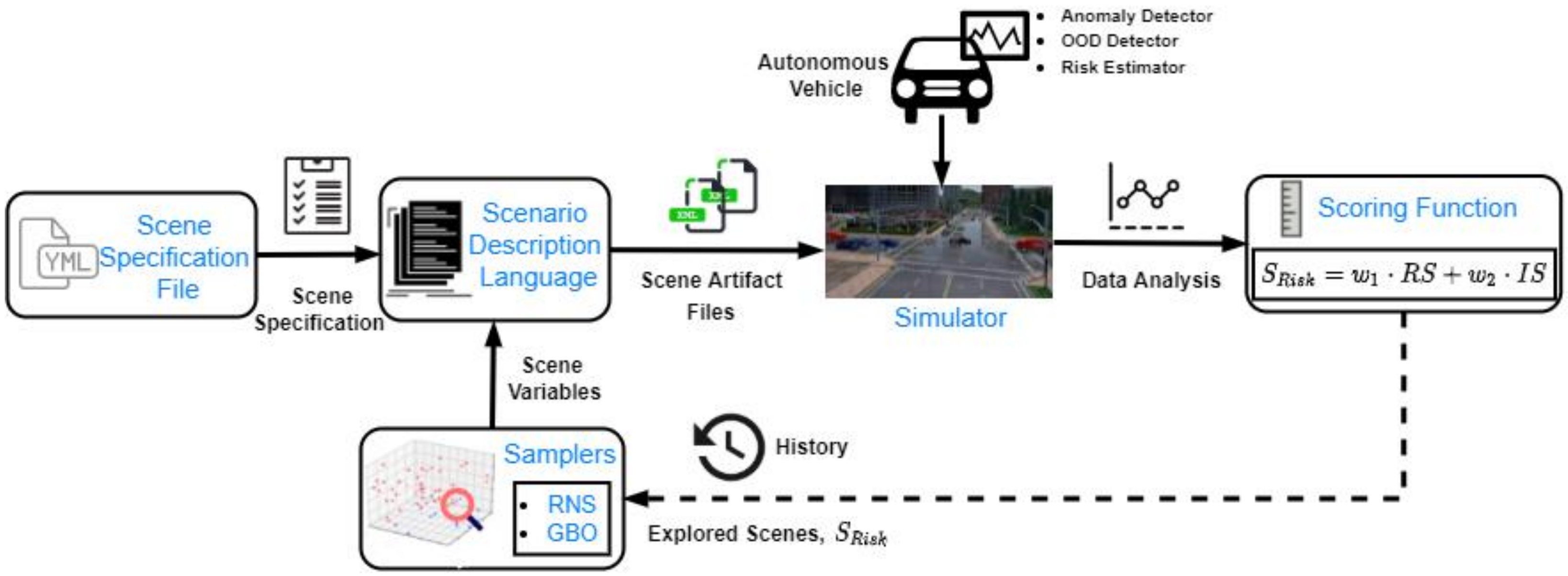




Key Results:

- Dynamic-weighted simplex had 60% fewer out-of-track occurrences at an average speeds of 0.4 m/s as compared to either LEC or OpenCV controller driven system

Automated Testing Framework



```
Name: World Description File
Scenario Description:
    town: 5 #Avaliable towns 5,6,7. A town can have N re
    regions: 5 #If regions are not available, the scenar
    weather:
      cloudiness: [0,100] # min:0, max:100
      precipitation: [0,100] # min:0, max:100
      precipitation_deposits: [0,100] # min:0, max:100
      sun_altitude_angle: [-90,90] # min:-90, max:90
      wind_intensity: false # min:0, max:100
      sun_azimuth_angle: false # min:0, max:100
      wetness: false # min:0, max:100
      fog_distance: false # min:0, max:100
      fog_density: false # min:0, max:100
    pedestrian_density: [0,2] #false -> no pedestian, tr
      #1: false
      #2: false
      #3: true
      #4: true
      #5: false
    traffic_density: [0,10] #false -> no traffic, true -
      #1: true
      #2: false
      #3: true
      #4: true
      #5: false
Constraints: #A constraint can be placed on the rate of
    pedestrian_density_delta: 1 #The pedestrian density
    traffic_density_delta: 2 #The traffic density will v
    weather_delta: 2 #The weather parameters will vary b
Initial Conditions: # A scenario can be strated from a s
    weather: #A value (not range) need to be specified f
      cloudiness: 0
      precipitation: 0
      precipitation_deposits: 0
      sun_altitude_angle: 45
      wind_intensity: 0
      sun_azimuth_angle: 0
      wetness: 0
      fog_distance: 0
      fog_density: 0
    region: 1 #scenario will start from region1
    traffic_density: 2
    pedestrian_density: 0
```

Domain-Specific Modeling language for test case generation

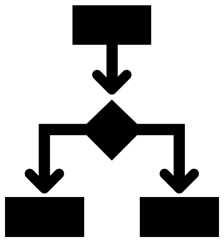
Github - https://github.com/scope-lab-vu/Risk-Aware-Scene-Generation-CPS

Summary

AI-Development → **Safety Assurance** → **Simulation & Testing**

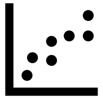
AI Robotics – Hardware & Software development

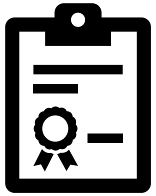
Decision Making under uncertainty



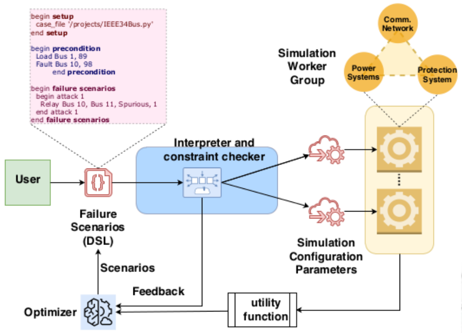
Assurance Case

Monitors
- Anomaly Detectors
- Runtime Verification
- Risk Estimator



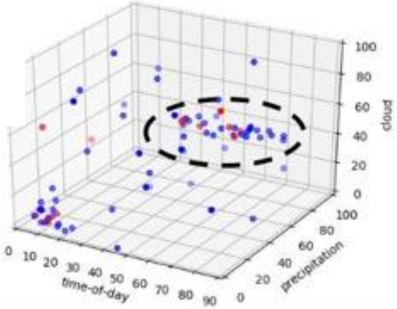


Safety Standards (ISO 26262, UL4600)

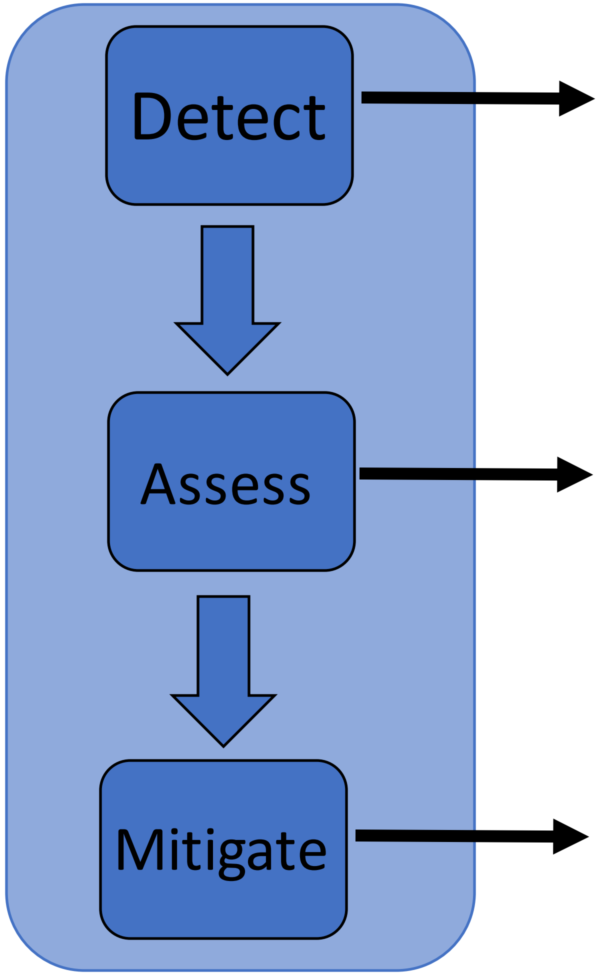


Domain-Specific Modeling Language

Design Space Exploration



Selected Publications



1. Sundar, V. K., **Ramakrishna, S.**, Rahiminasab, Z., Easwaran, A., & Dubey, A. (2020, May). "Out-of-distribution detection in multi-label datasets using latent space of β-VAE". In *2020 IEEE Security and Privacy Workshops (SPW)*
2. **Ramakrishna, S.**, Rahiminasab, Z., Easwaran, A., & Dubey, A. (2020, September). "Efficient Multi-Class Out-of-Distribution Reasoning for Perception Based Networks: Work-in-Progress." In *2020 International Conference on Embedded Software (EMSOFT)*
3. **Ramakrishna, S.**, Rahiminasab, Z., Karsai, G., Easwaran, A., & Dubey, A. (2021). "Efficient Out-of-Distribution Detection Using Latent Space of β-VAE for Cyber-Physical Systems." in TCPS 2020
4. Burruss, M., **Ramakrishna, S.**, & Dubey, A. (2021). "Deep-RBF Networks for Anomaly Detection in Automotive Cyber-Physical Systems." In SmartComp 2021

1. Hartsell, C.*, **Ramakrishna, S.***, Dubey, A., Stojcsics, D., Mahadevan, N., & Karsai, G. (2021). "ReSonAte: A Runtime Risk Assessment Framework for Autonomous Systems". In SEAMS 2021
2. **Ramakrishna, S.**, Luo, B., Barve, Y., Karsai, G., & Dubey, A. (2021). "Risk-Aware Scene Sampling for Dynamic Assurance of Autonomous Systems" submitted to ICAA 2021
3. **Ramakrishna, S.**, Hartsell, C., Dubey, A., Pal, P., & Karsai, G. (2020). "A Methodology for Automating Assurance Case Generation". In TMCE 2020

1. **Ramakrishna, S.**, Dubey, A., Burruss, M. P., Hartsell, C., Mahadevan, N., Nannapaneni, S., ... & Karsai, G. (2019, May). "Augmenting learning components for safety in resource constrained autonomous robots." In *2019 IEEE 22nd International Symposium on Real-Time Distributed Computing (ISORC)*
2. **Ramakrishna, S.**, Harstell, C., Burruss, M. P., Karsai, G., & Dubey, A. (2020). "Dynamic-weighted simplex strategy for learning enabled cyber physical systems." *Journal of systems architecture*
3. Burruss, M. P., **Ramakrishna, S.**, Karsai, G., & Dubey, A. (2019, May). "Deepnncar: A testbed for deploying and testing middleware frameworks for autonomous robots." In *2019 IEEE 22nd International Symposium on Real-Time Distributed Computing (ISORC)*