



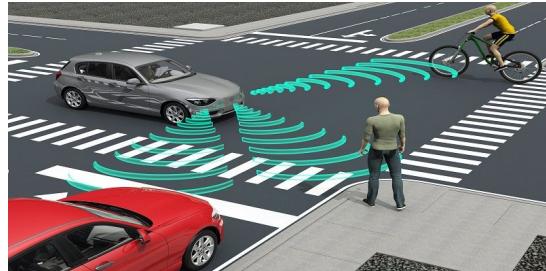
Safety Evaluation for Self-Driving Cars (and Other Intelligent Systems)

Mansur M. Arief, Ph.D.

Postdoctoral Scholar, Stanford Intelligent Systems Lab

Email: mansur.arief@stanford.edu | Website: www.mansurarief.github.io

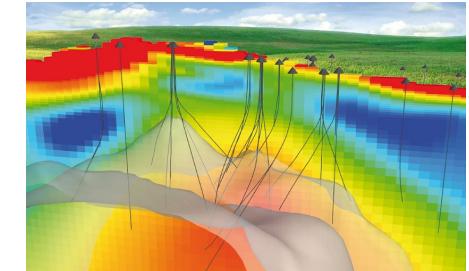
High-stake decisions are clouded with uncertainty



Autonomy stack

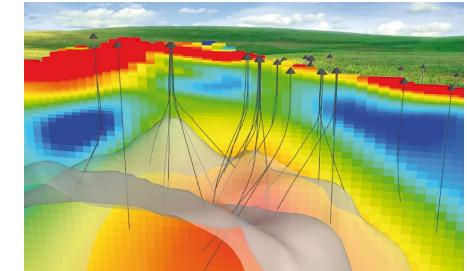
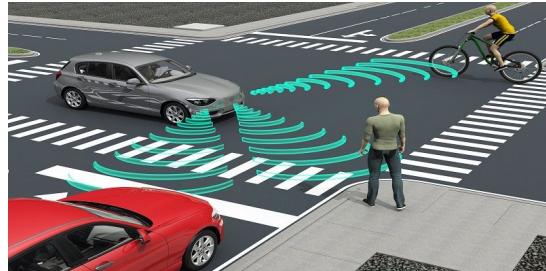


Mineral exploration



Renewable energy

Effective support systems plays a crucial role



Preventing accidents



Most-efficient operations



Sustainable operations

AI autonomy is here...

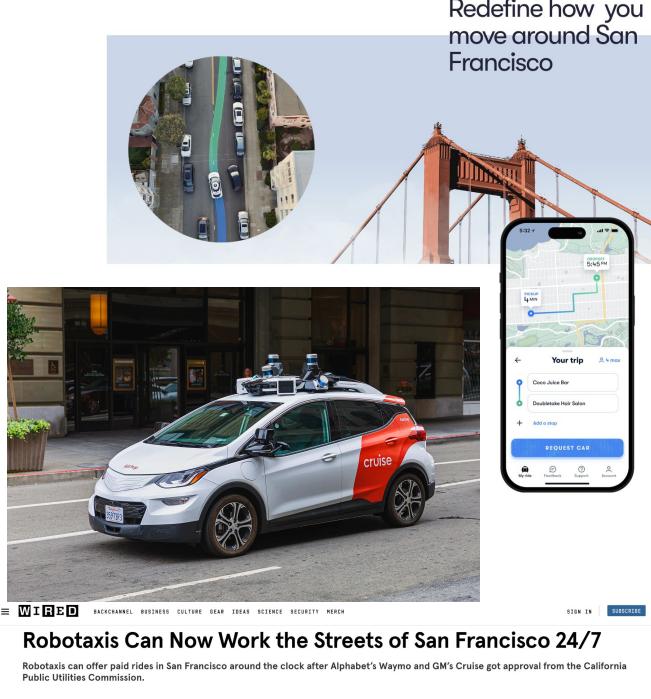
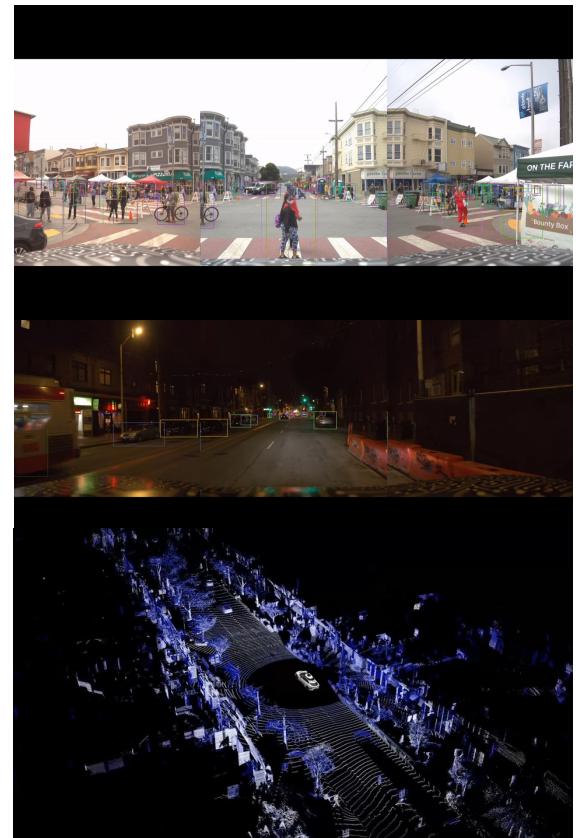


Photo I took two weeks ago, in an Uber ride

AI autonomy is here...

How self-driving cars “see” their environments

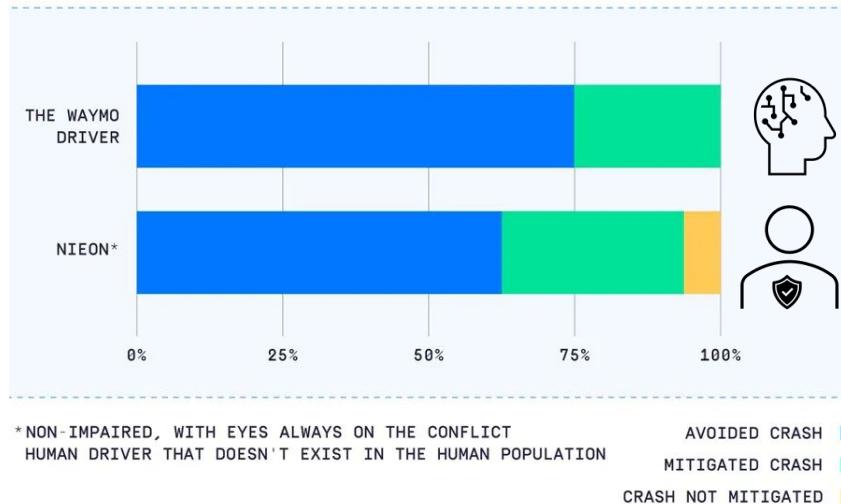


AI autonomy is here...

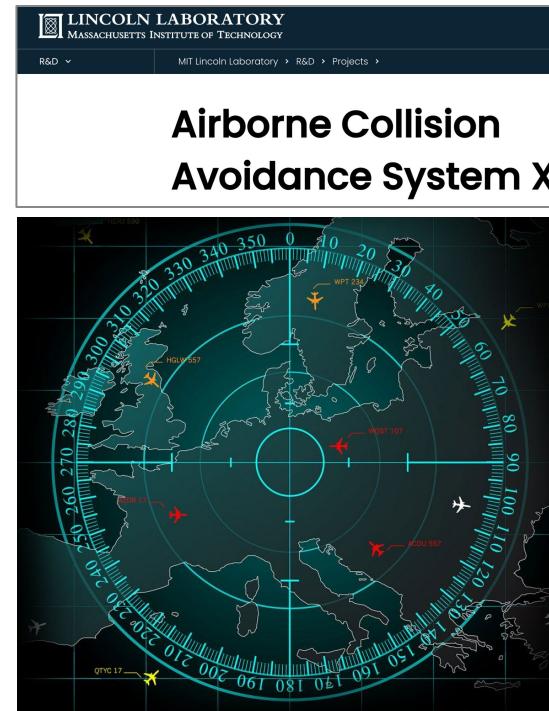


AI autonomy is here... and becomes safer everyday

The Waymo Driver's collision avoidance performance in simulated tests



Source: <https://www.theverge.com/2022/9/29/23377219/waymo-av-safety-study-response-time-crash-avoidance>,
<https://waymo.com/waymo-one-san-francisco/>



A next-generation collision avoidance system will help pilots and unmanned aircraft safely navigate the airspace.

Is it safe enough?

Google cruise self driving san francisco

All News Images Videos Maps More Tools

About 1,100 results (0.35 seconds)

 **Boing Boing**

[Watch someone in San Francisco smashing a Cruise autonomous car with a hammer \(video\)](#)

In a scene that looks like a TV afterschool special about cyberpunks, a person dressed in black attempted to damage a Cruise autonomous...

1 day ago



 **The Guardian**

[Self-driving car blocking road 'delayed patient care', San Francisco officials say](#)

Cruise, the robotaxi firm, denies the city's claims its vehicle blocked ambulance which resulted in injured person's death.

1 week ago



 **The New York Times**

[Driverless Taxis Blocked Ambulance in Fatal Accident, San Francisco Fire Dept. Says](#)

Two Cruise taxis delayed an ambulance carrying a car accident victim to a hospital, a department report said. The company said it was not at...

2 weeks ago



 **Washington Post**

[Cruise CEO calls criticism of San Francisco robotaxis 'sensationalism'](#)

SAN FRANCISCO — Residents and city officials here are increasingly fed up with the self-driving cars that have blanketed the city...

1 week ago



 **NPR**

[Protesters stop Waymo and Cruise self-driving cars with only a traffic cone](#)

Self-driving cars have flooded San Francisco's streets and not everyone is happy



Google waymo self driving san francisco

All News Images Videos Maps More Tools

 **The San Francisco Standard**

[San Franciscans Are Having Sex in Robotaxis, and Nobody Is Talking About It](#)

A little-known 2018 study said the spread of self-driving cars was likely to mean more sex on the road. San Franciscans are making it...

1 month ago



 **WIRED**

[Uber and Lyft Drivers Have Some Advice for Autonomous Vehicles Set to Swarm the Streets](#)

San Francisco ride-hail drivers are about to share the roads with robot competitors. They say that the self-driving cabs need to work on...

1 month ago



 **Vox**

[San Francisco's self-driving taxi experiment with Google and GM is causing some chaos](#)

Self-driving taxis are ferrying passengers across San Francisco and Phoenix, and they could be coming to a street near you very soon.

4 weeks ago



 **Electrek**

[Waymo starts taking fares for SF robotaxis after state approval](#)

Waymo received final approval to operate its driverless taxis in California last week and will start charging for its driverless robotaxi...

1 month ago



 **The Guardian**

[Robotaxi breakdowns cause mayhem in San Francisco days after expansion vote](#)

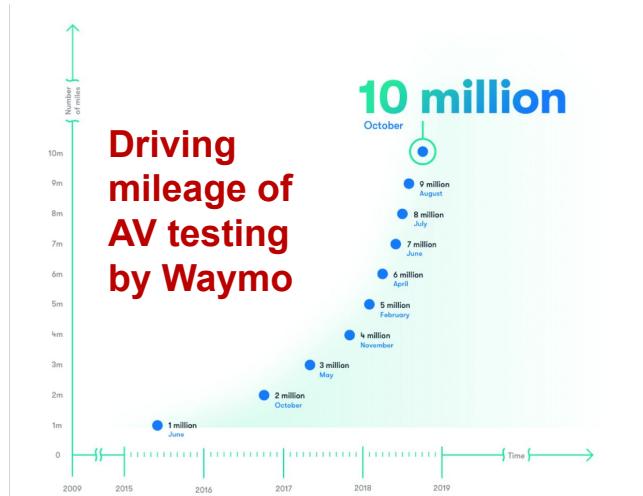
Public utilities commission had allowed Google's Waymo and General Motors' Cruise to operate all day in a vote on Thursday.

1 month ago



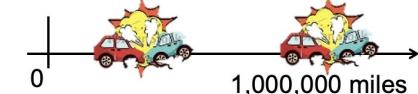
Safe AI is hard to evaluate

- If the **failure rate is μ** , then on average we need **$1/\mu$ samples** to observe the first failure (geometric distribution).
- Hence, **smaller μ requires larger sample size.**



Source: [Waymo](#)

Main reason:

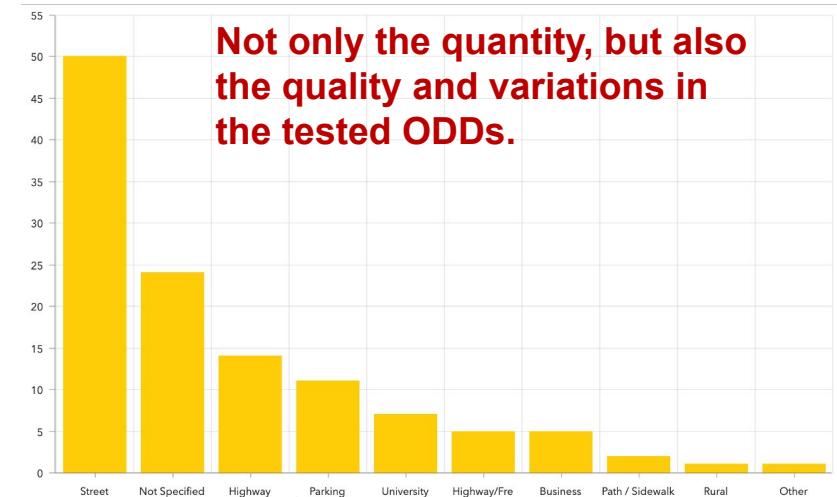


Crashes happen extremely **rarely** (NHTSA, 2019).

Safe AI is hard to evaluate



Source: <https://www.nhtsa.gov/automated-vehicle-test-tracking-tool>



Not only the quantity, but also the quality and variations in the tested ODDs.

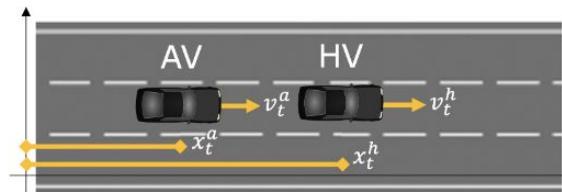
- The huge testing burden results in **costly, lengthy, and risky** on-road deployments.¹

¹Arief, Mansur, Glynn, P., & Zhao, D. An accelerated approach to safely and efficiently test pre-production autonomous vehicles on public streets. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 2006-2011). IEEE, 2018.

Inefficiency remains an issue in simulations



AV Simple PI Controller (SAE Level 2):



$$a_t^a = a_{t-1}^a + K_p(e_t^{thw} - e_{t-1}^{thw}) + K_i(e_t^{thw} + e_{t-1}^{thw})T_s/2$$

where a_t^a : AV acceleration at time t

e_t^{thw} : target and realization time headway error at time t

K_p, K_i : P and I gain, respectively

T_s : simulation frequency

- Naturalistic simulation takes up to **a month of runtime** to estimate $\mu = 2 \times 10^{-5}$

Inefficiency remains an issue in simulations

AV Perception Algorithm (YOLOv5)



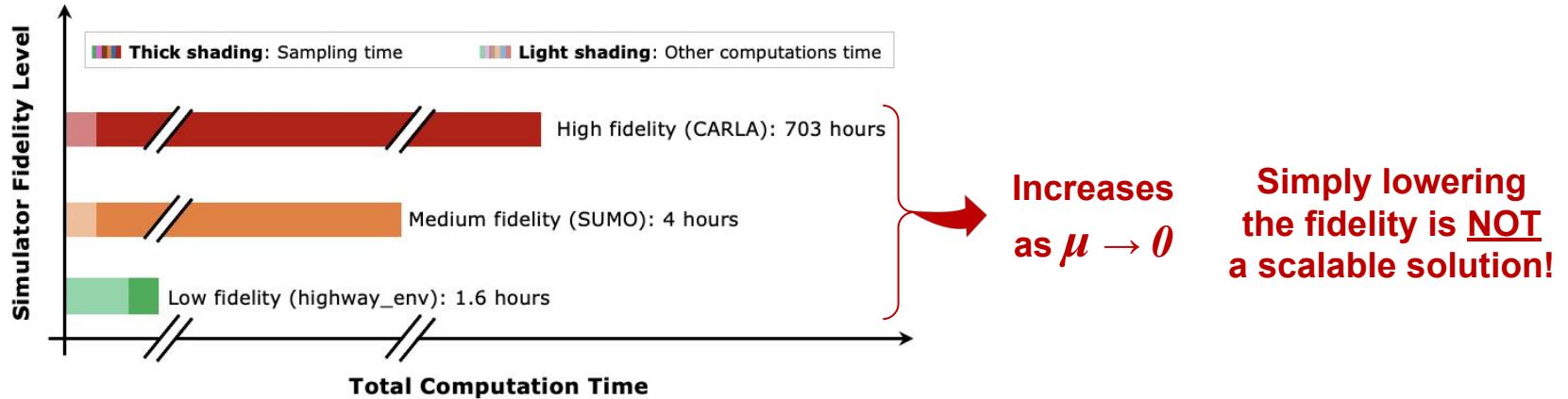
Normal cases



Extremely rare (1 in 1 million simulation)

- May take **3 months (estimated) runtime** to estimate smaller $\mu = 1 \times 10^{-6}$

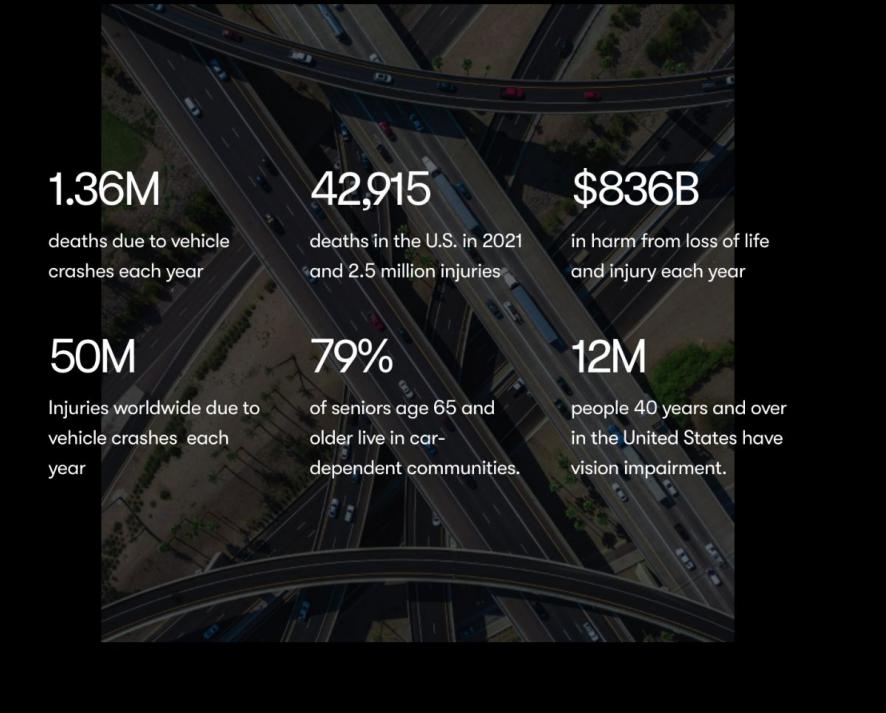
Inefficiency remains an issue in simulations



Why must we care?

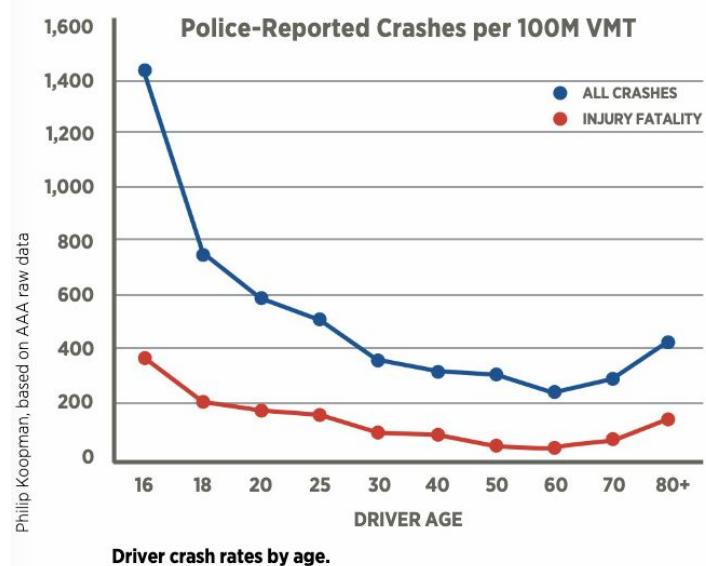
Because Safety
is Urgent™

Autonomous Driving Technology
Can Save Lives and Improve
Mobility



How to certify safety?

- **Main focus:** probabilistic evaluation
 - compares to **baseline** and standards
 - allows **risk mitigation** strategies
 - more **practical** and aligns with engineering continuous improvement
 - ISO 26262, ISO 21448
 - UL4600, SAE J3018



Source: [SAE Update p. 31](#)

Can we just list all test cases? (EuroNCAP)

- No, because AI is “smart” and trainable
- We recently published a paper on generating safety-critical test cases



Mansur Maturidi Arief

FOLLOW

[Stanford University](#)

Verified email at stanford.edu - [Homepage](#)

Rare-event simulation importance sampling autonomous vehicle safety analysis sustainability

TITLE	CITED BY	YEAR
A survey on safety-critical driving scenario generation—A methodological perspective W Ding, C Xu, M Arief, H Lin, B Li, D Zhao IEEE Transactions on Intelligent Transportation Systems	50	2023

but it's not enough. We want to get “unbiased” safety estimate

Scalable and certifiable evaluation algorithms

- **Objective:** Develop algorithms that can deal with
 - extreme rarity and high-dimensional inputs
- **Requirements:**
 - efficiency guarantee and efficient computation
- **Proposed algorithms:**
 - Deep IS: Deep Importance Sampling¹
 - Deep-PrAE: Deep Probabilistic Accelerated Evaluation²
 - CERTIFY: Computationally Efficient and Robust Evaluation of Safety³

¹Arief, Mansur, Zhepeng Cen, Zhenyuan Liu, Zhiyuan Huang, Bo Li, Henry Lam, and Ding Zhao. "Certifiable Evaluation for Autonomous Vehicle Perception Systems Using Deep Importance Sampling (Deep IS)." In *Proceedings of the 2022 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022. [\[Link\]](#)

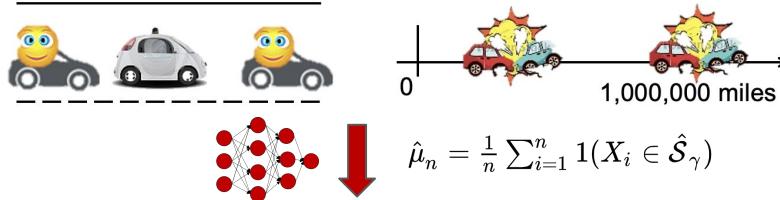
²Arief, Mansur, Zhiyuan Huang, Guru Koushik Senthil Kumar, Yuanlu Bai, Shengyi He, Wenhao Ding, Henry Lam, and Ding Zhao. "Deep Probabilistic Accelerated Evaluation: A Certifiable Rare-Event Simulation Methodology for Black-Box Autonomy." In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 2021. [\[Link\]](#)

³Arief, Mansur, Zhepeng Cen, Huan Zhang, Henry Lam, and Ding Zhao. "CERTIFY: Computationally Efficient Rare-failure Certification of Autonomous Vehicles." *Under review for IEEE T-IV.* [\[Link\]](#)

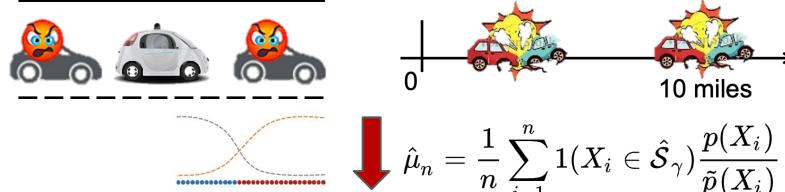
Statistical Workhorse: Importance Sampling

- High-level idea

Naturalistic driving conditions:



Aggressive driving conditions:



Unbiased result

Key steps:

- Start with normal driving
- Learn the statistical model
- Bias the statistics toward more aggressive driving
- Use importance weights to obtain unbiased result
- Return unbiased statistics

Deep IS: Unbiased, given an accurate approximation

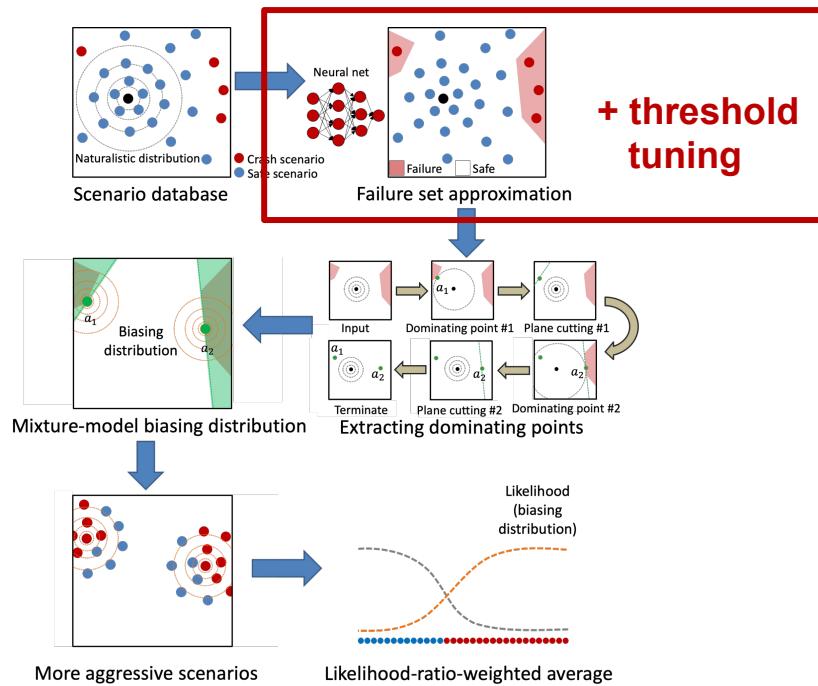
- Suppose NN gives a set approximation $\hat{\mathcal{S}}_\gamma \approx \mathcal{S}_\gamma$ (the true failure rate) after training with n_1 samples. We have, with $n_2 = n - n_1$ samples,

$$\begin{aligned}
 \mathbb{E}_{X \sim p}[\hat{\mu}_n] &= \mathbb{E} \left[\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbb{1}(X_i \in \hat{\mathcal{S}}_\gamma) L(X_i) \right] \\
 &= \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbb{E} \left[\mathbb{1}(X_i \in \hat{\mathcal{S}}_\gamma) \frac{\phi(\tilde{X}_i; \lambda, \Sigma)}{\sum_{a \in \hat{\mathcal{A}}_\gamma} w_a \phi(\tilde{X}_i; a, \Sigma)} \right] \\
 &= \frac{1}{n_2} \sum_{i=1}^{n_2} \int_{\mathbb{R}^d} \mathbb{1}(X_i \in \hat{\mathcal{S}}_\gamma) \frac{\phi(\tilde{X}_i; \lambda, \Sigma)}{\sum_{a \in \hat{\mathcal{A}}_\gamma} w_a \phi(\tilde{X}_i; a, \Sigma)} \sum_{a \in \hat{\mathcal{A}}_\gamma} w_a \phi(\tilde{X}_i; a, \Sigma) dX_i \\
 &= \frac{1}{n_2} \sum_{i=1}^{n_2} \int_{\mathbb{R}^d} \mathbb{1}(X_i \in \hat{\mathcal{S}}_\gamma) \phi(\tilde{X}_i; \lambda, \Sigma) dX_i \\
 &= \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbb{E}_{X \sim p} \mathbb{1}(X_i \in \hat{\mathcal{S}}_\gamma) \\
 &\approx \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbb{E}_{X \sim p} \mathbb{1}(X_i \in \mathcal{S}_\gamma) \\
 &= \boxed{\mu}.
 \end{aligned}$$

If we have accurate deep learning classifier,
we have provable unbiased results!

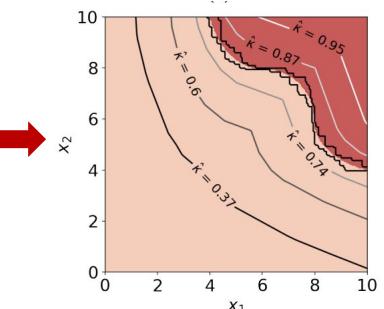
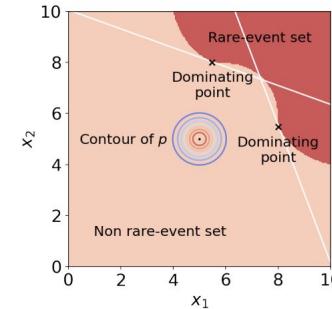
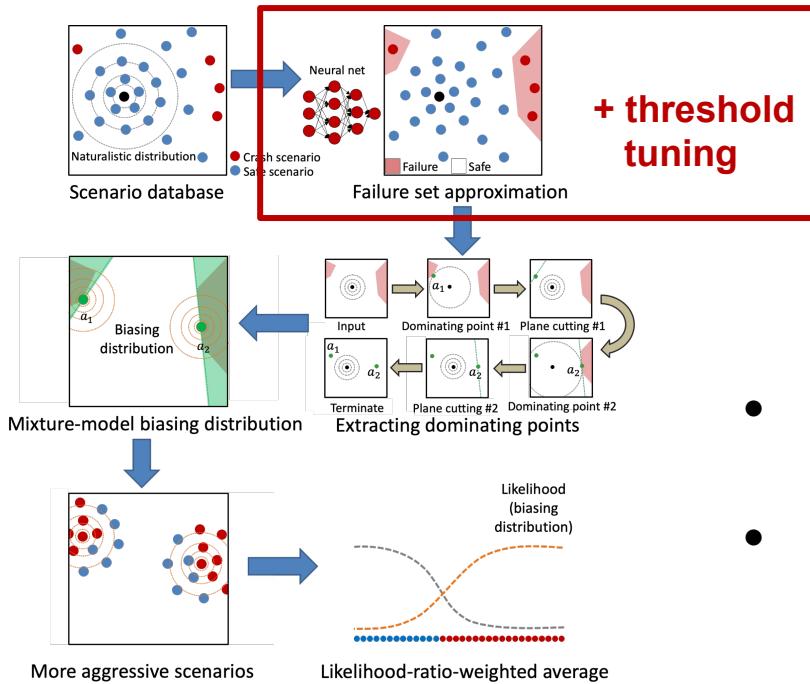
Deep-PrAE: Deep Probabilistic Accelerated Evaluation

- What if we have an error, can we prove efficiency? Yes, a conservative one!



Deep-PrAE: Deep Probabilistic Accelerated Evaluation

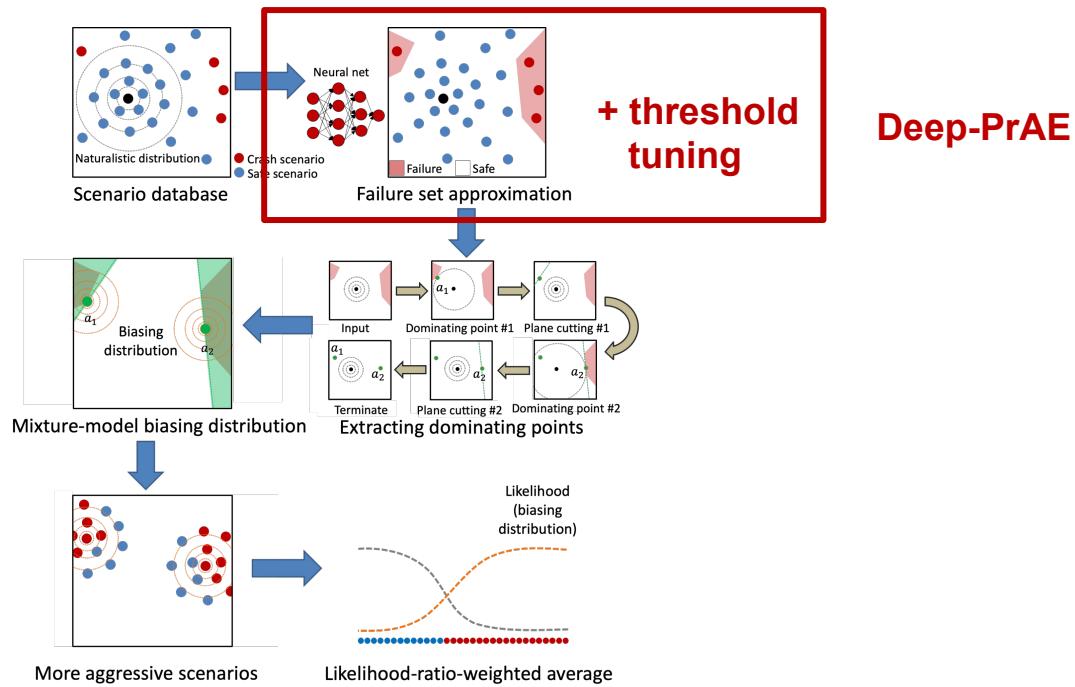
- What if we have an error, can we prove efficiency? Yes, a conservative one!



- With this, we have an upper-bound for the failure probability, but it is still useful for safety evaluation
- If an upper bound for something is below some value, then its true value must be below it too

CERTIFY: Computationally Efficient and Robust Evaluation of Safety

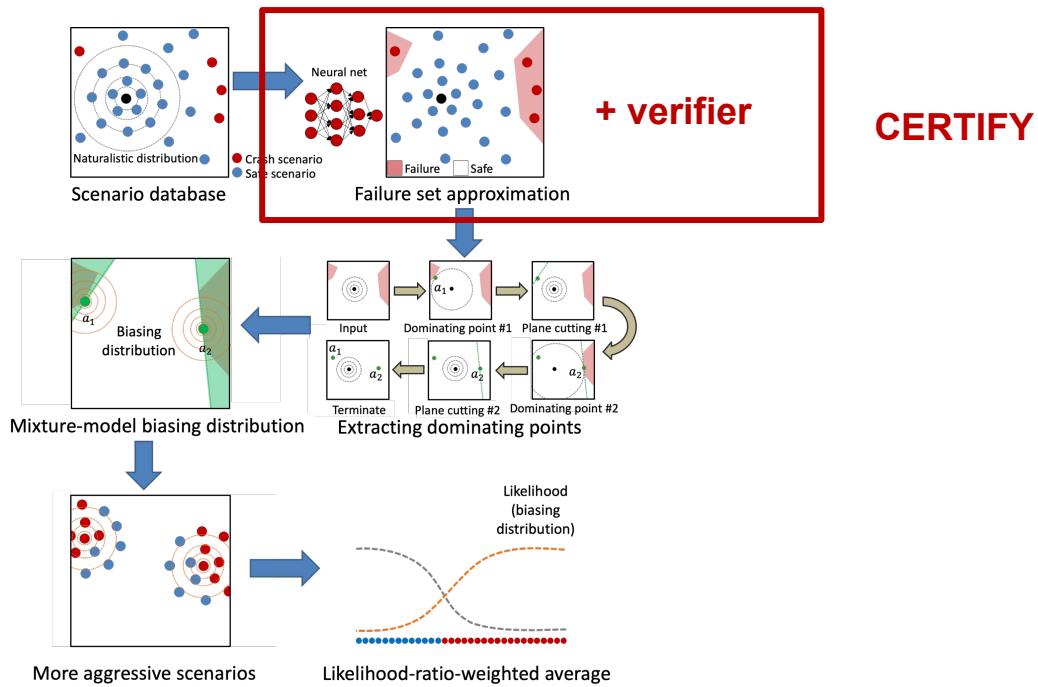
- Main idea: Can we speed up? Yes, by simplifying neural network constraints!



³Arief, Mansur, Zhepeng Cen, Huan Zhang, Henry Lam, and Ding Zhao. "CERTIFY: Computationally Efficient Rare-failure Certification of Autonomous Vehicles." *Under review*.

CERTIFY: Computationally Efficient and Robust Evaluation of Safety

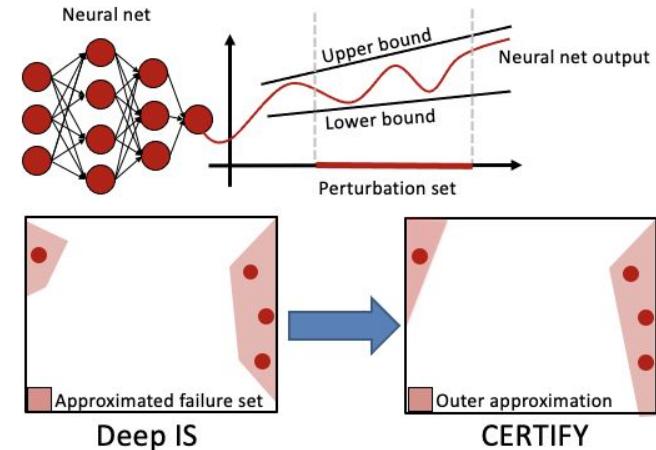
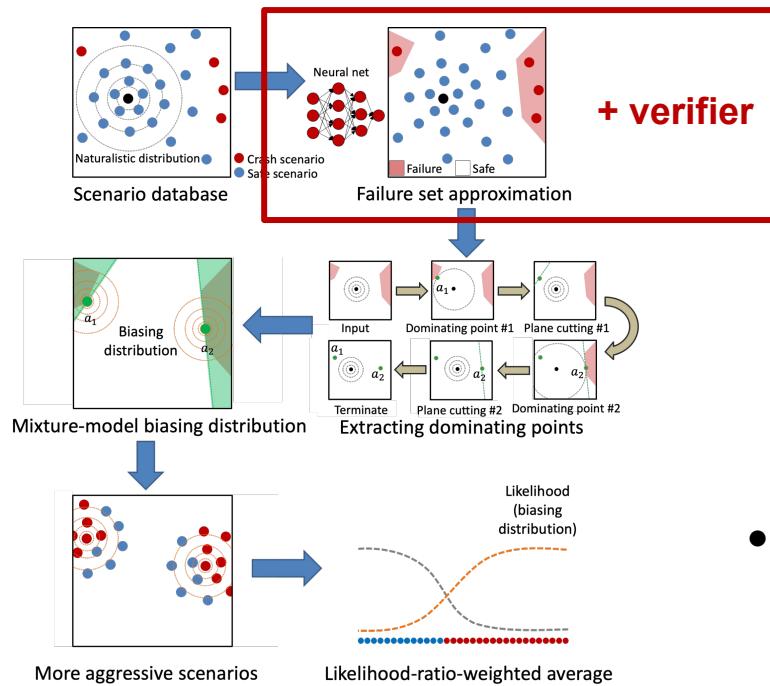
- Main idea: Can we speed up? Yes, by simplifying neural network constraints!



³Arief, Mansur, Zhepeng Cen, Huan Zhang, Henry Lam, and Ding Zhao. "CERTIFY: Computationally Efficient Rare-failure Certification of Autonomous Vehicles." *Under review*.

CERTIFY: Computationally Efficient and Robust Evaluation of Safety

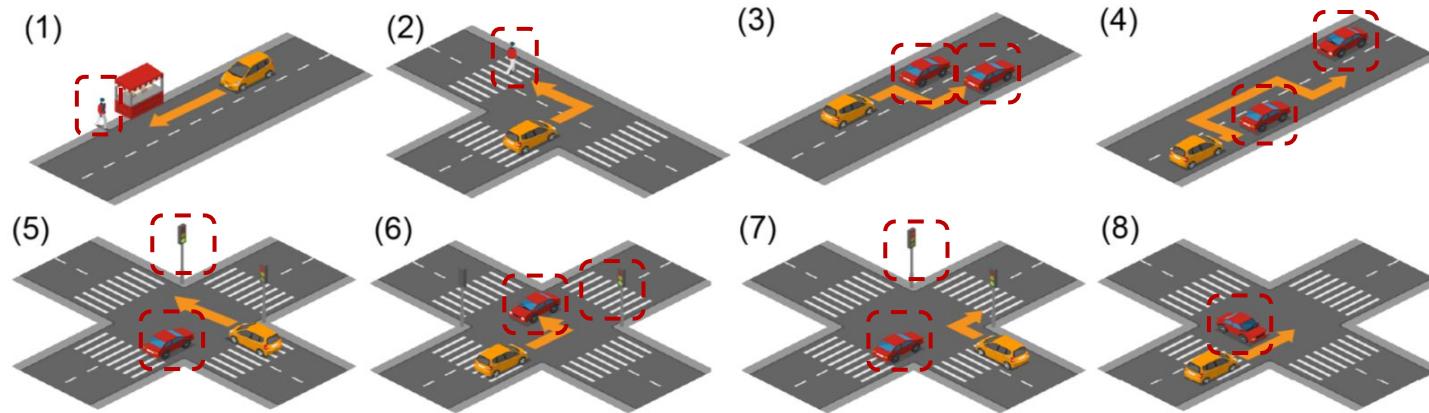
- Main idea: Can we speed up? Yes, by simplifying neural network constraints!



- With this, we have a valid upper-bound that is fast to compute, but might be more conservative

Benchmarking settings

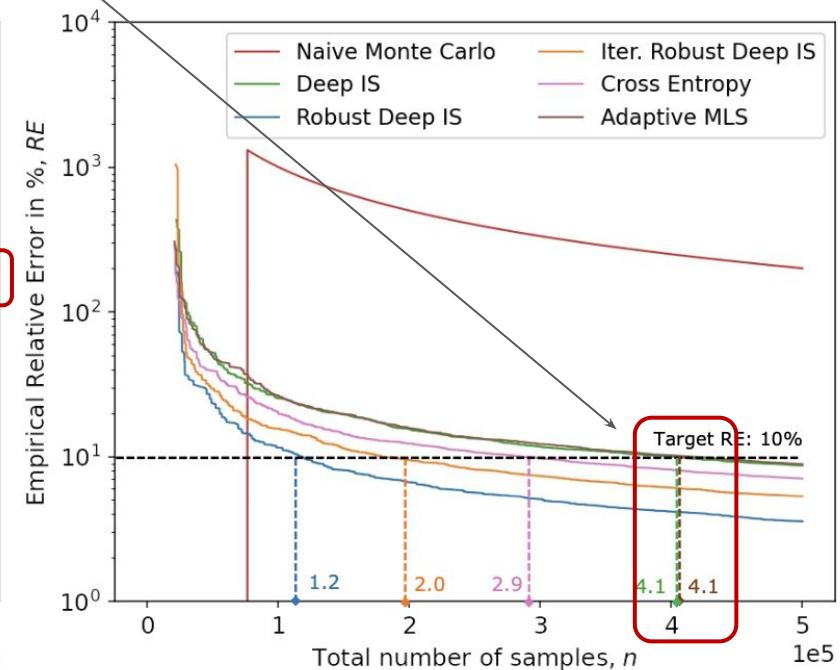
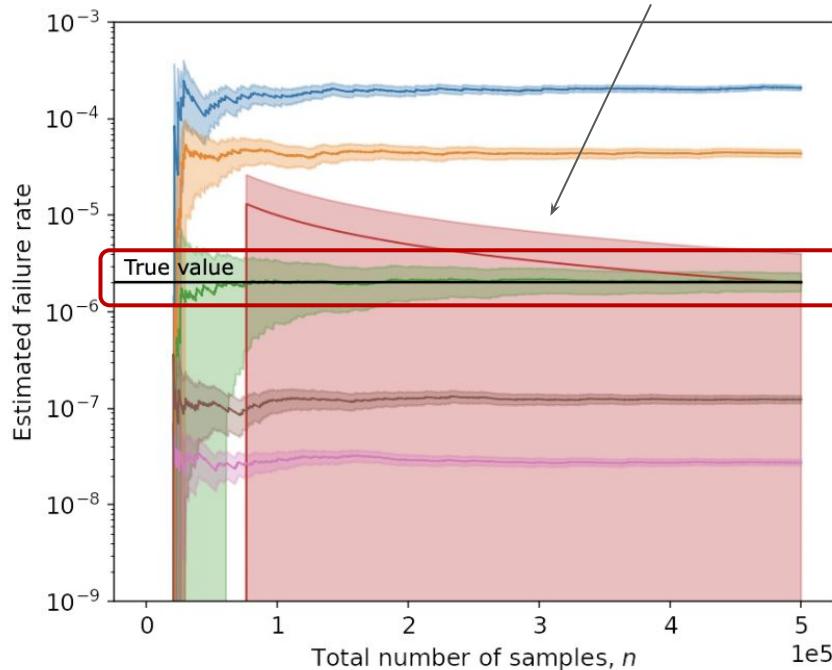
- **Driving scenarios:** Eight driving scenarios as defined in SafeBench [1].



(1) Straight Obstacle, (2) Turning Obstacle, (3) Lane Changing, (4) Vehicle Passing,
(5) Red-light Running, (6) Unprotected Left-turn, (7) Right-turn, and (8) Crossing Negotiation.

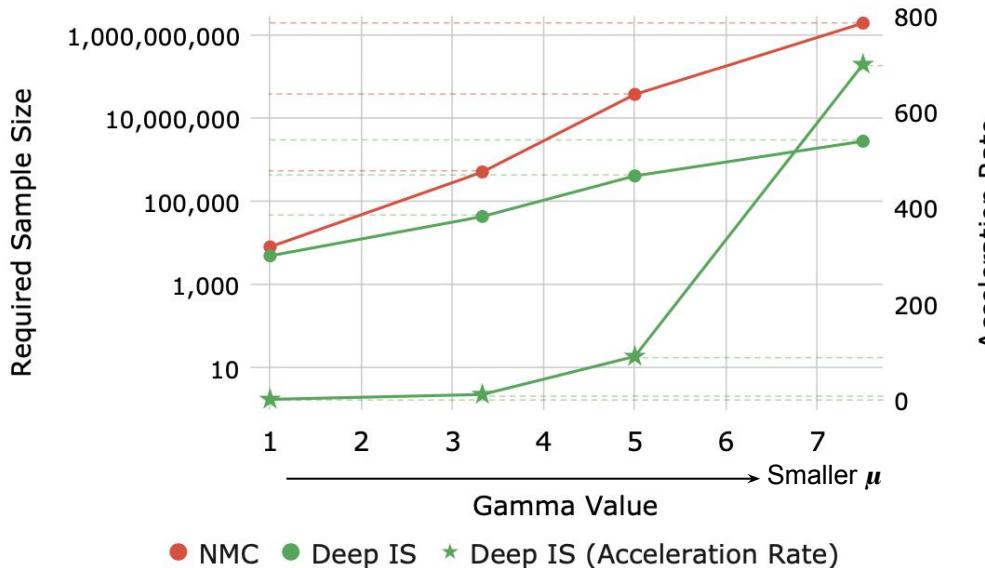
Deep IS numerical experiments

- Main result: Deep IS is unbiased and sample-efficient**



Deep IS numerical experiments

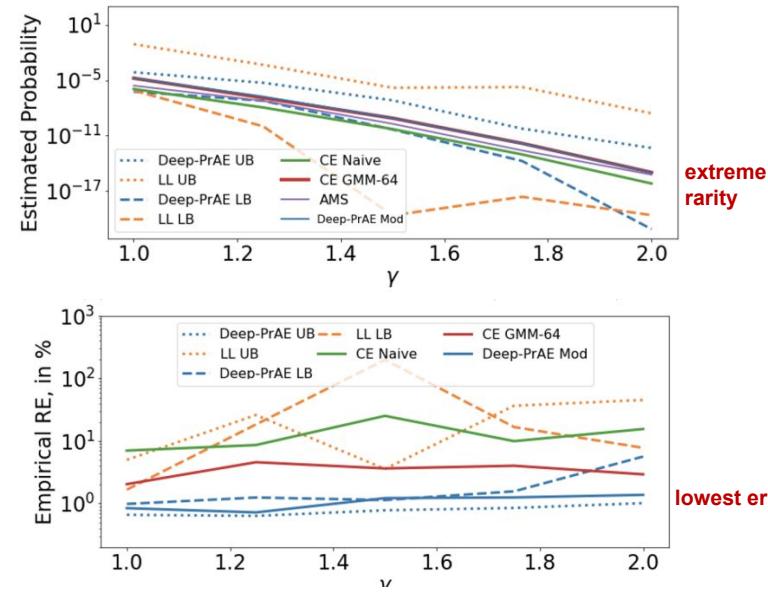
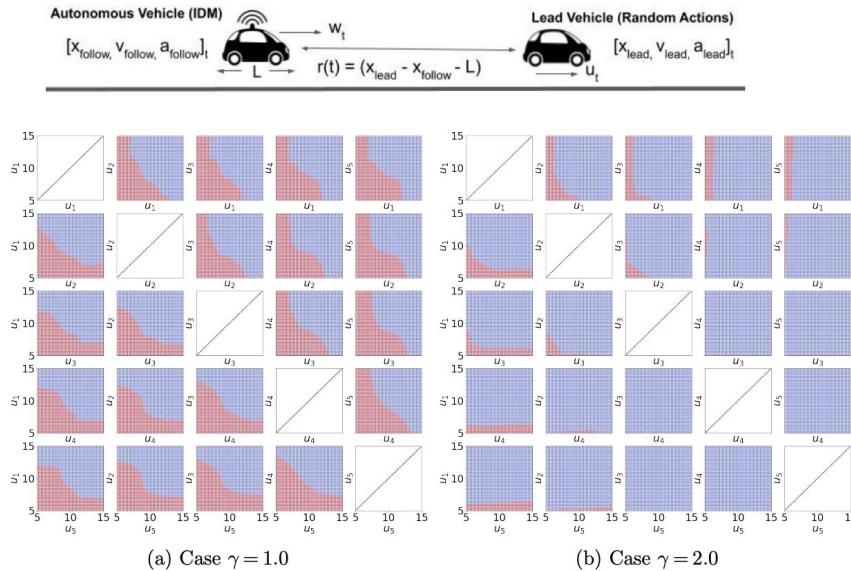
- Main result: (Empirically) Handles rarity well with efficiency boost!



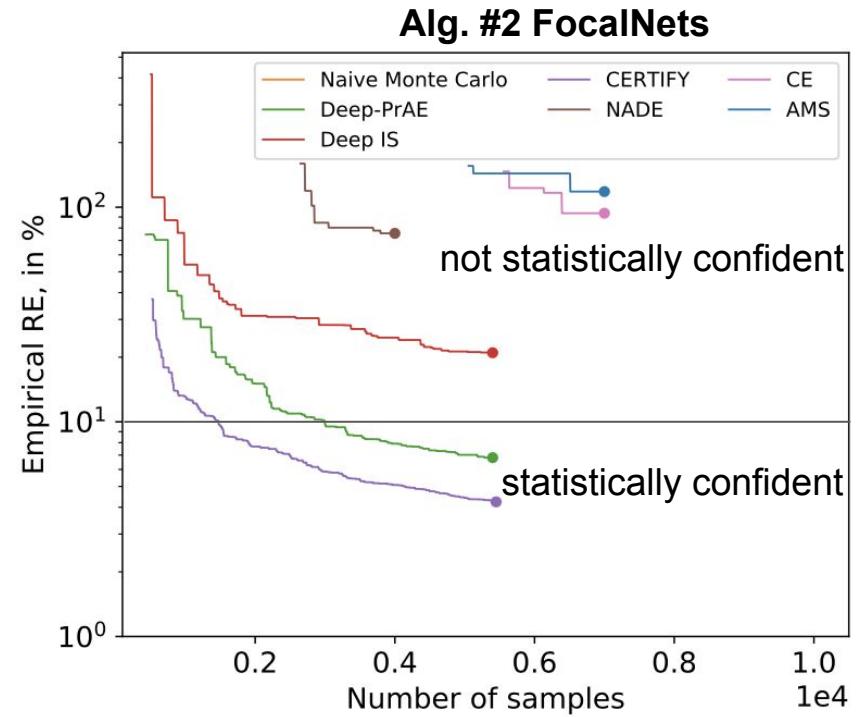
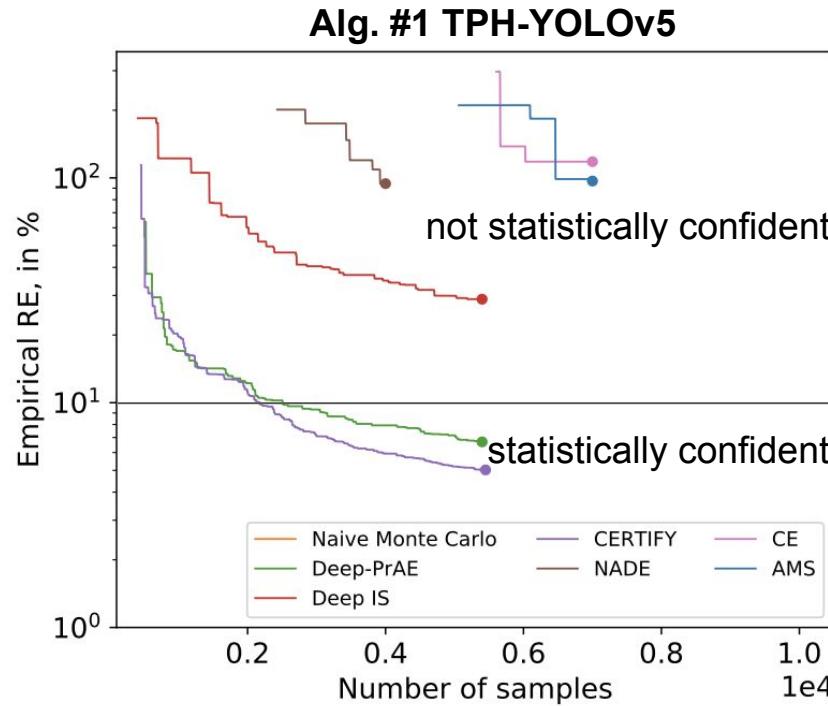
¹Arief, Mansur, Zhepeng Cen, Zhenyuan Liu, Zhiyuan Huang, Bo Li, Henry Lam, and Ding Zhao. "Certifiable Evaluation for Autonomous Vehicle Perception Systems Using Deep Importance Sampling (Deep IS)." In *Proceedings of the 2022 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022. [[Link](#)]

Deep-PrAE numerical experiments

- Autonomy evaluation example: Deep-PrAE efficiency dominates.



CERTIFY numerical experiments



- Only Deep-PrAE and CERTIFY achieve high confidence (<10% RE).

Summary

- Testing complex (blackbox) AI with rare failure cases is challenging
- The “intelligence” notion of AI makes conventional testing approaches ineffective
- Other domains with rare but critical failure events can be evaluated similarly
- In the context of safety, the notion of statistical confidence should be emphasized (which makes safety evaluation more challenging)

¹Arief, Mansur, Zhepeng Cen, Zhenyuan Liu, Zhiyuan Huang, Bo Li, Henry Lam, and Ding Zhao. "Certifiable Evaluation for Autonomous Vehicle Perception Systems Using Deep Importance Sampling (Deep IS)." In *Proceedings of the 2022 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022. [[Link](#)]

Collaborations with Institutions in Indonesia

- **DSA-POMDP**
 - How can we optimize stroke patient treatment as a sequential decision-making framework, determining when to perform surgery, MRA, DSA, etc. with limited prior information?
 - With 2 MD graduates, from UI and UIN Syarif Hidayatullah
- **Robust Design of EV Charging Stations in Indonesia**
 - We are using variance reduction technique into simulation-based optimization to determine infrastructure for EVs under outage uncertainty
 - With ITS and Carnegie Mellon
- **Intelligent Decision Support System for Agro-Products and Sustainable Mining**
 - How to best integrate IoT, smart devices, and deep learning to reduce food loss and maximize total values - with IPB and Carnegie Mellon
 - How to design proactive community development enterprise and ensure sustainable ecosystems during and post-mining operations - with MineralX and Unhas

Collaboration with Interdisciplinary Scholars

- **IndoSTEELERS:** Indonesian Scholars Thriving in Excellence in Education, Learning, and Research
- Scholars (professors, practitioners, postdocs, graduate students) in
 - law
 - music and society
 - education
 - health informatics
 - public policy
 - intelligent systems, including AD, LLM, and robotics
- Open to collaborate for relevant researches in Indonesian context



Research Areas

VERIFICATION & VALIDATION

Development of efficient verification and validation algorithms for autonomous systems.

1. Ding, Wenhao, Chejian Xu, Mansur Arief, Haohong Lin, Bo Li, Ding Zhao. "A Survey on Safety-Critical Driving Scenario Generation—A Methodological Perspective." *T-ITS*, 2023.
<https://ieeexplore.ieee.org/abstract/document/10089194>
2. Arief, Mansur. "Certifiable Evaluation for Safe Intelligent Autonomy." *Carnegie Mellon University*, 2023.
<https://www.proquest.com/openview/45f55565d4810a203cc28fc50dd878a6>
3. Arief, Mansur, Zhepeng Cen, Zhenyuan Liu, Zhiyuan Huang, Bo Li, Henry Lam, and Ding Zhao. "Certifiable Evaluation for Autonomous Vehicle Perception Systems Using Deep Importance Sampling (Deep IS)." *ITSC*, 2022.
<https://ieeexplore.ieee.org/abstract/document/9922202>
4. Arief, Mansur, Yuanlu Bai, Wenhao Ding, Shengyi He, Zhiyuan Huang, Henry Lam, and Ding Zhao. "Certifiable Deep Importance Sampling for Rare-Event Simulation of Black-Box Systems." *Under Review*.
<https://arxiv.org/abs/2111.02204>
5. Arief, Mansur, Zhiyuan Huang, Guru Koushik Senthil Kumar, Yuanlu Bai, Shengyi He, Wenhao Ding, Henry Lam, and Ding Zhao. "Deep Probabilistic Accelerated Evaluation: A Certifiable Rare-Event Simulation Methodology for Black-Box Autonomy." *AISTATS*, 2021.
<https://proceedings.mlr.press/v130/arief21a/arief21a.pdf>
6. Chen, Rui, Mansur Arief, Weiyang Zhang, and Ding Zhao. "How to Evaluate Proving Grounds for Self-Driving? A Quantitative Approach." *T-ITS*, 2020.
<https://ieeexplore.ieee.org/document/9094370>
7. Huang, Zhiyuan, Mansur Arief, Henry Lam, and Ding Zhao. "Evaluation Uncertainty in Data-Driven Self-Driving Testing." *ITSC*, 2019.
<https://ieeexplore.ieee.org/abstract/document/8917406>
8. Arief, Mansur, Peter Glynn, and Ding Zhao. "An Accelerated Approach to Safely and Efficiently Test Pre-production Autonomous Vehicles on Public Streets." *ITSC*, 2018.
<https://ieeexplore.ieee.org/document/9094370>

Research Areas

AUTONOMOUS DRIVING PERCEPTION

Works that design robust perception systems for autonomous driving applications.

1. Abdussyukur, Hafizh, Mahmud Dwi Sulistyo, Ema Rachmawati, Mansur Arief, Gamma Kosala. "Semantic Segmentation for Identifying Road Surface Damages Using Lightweight Encoder-Decoder Network." *ICACNIS*, 2022.
<https://ieeexplore.ieee.org/abstract/document/10056030>
2. Arief, Hasan Asy'ari, Mansur Arief, Guilin Zhang, Zuxin Liu, Manoj Bhat, Ulf Geir Indahl, Håvard Tveite, and Ding Zhao. "SAnE: Smart Annotation and Evaluation Tools for Point Cloud Data." *IEEE Access*, 2020.
<https://ieeexplore.ieee.org/iel7/6287639/8948470/09143095.pdf>
3. Liu, Zuxin, Mansur Arief, and Ding Zhao. "Where Should We Place LiDARs on the Autonomous Vehicle? An Optimal Design Approach." *ICRA*, 2019.
<https://ieeexplore.ieee.org/document/8793619>
4. Arief, Hasan Asy'ari, Mansur Arief, Manoj Bhat, Ulf Geir Indahl, Håvard Tveite, and Ding Zhao. "Density-Adaptive Sampling for Heterogeneous Point Cloud Object Segmentation in Autonomous Vehicle Applications." *CVPR Workshops*, 2019.
https://openaccess.thecvf.com/content_CVPRW_2019/papers/UG2+20Prize%20Challenge/Arief_Density-Adaptive_Sampling_for_Heterogeneous_Point_Cloud_Object_Segmentation_in_Autonomous_CVPRW_2019_paper.pdf

Research Areas

EV & INFRASTRUCTURE

Studies focused on vehicle electrification and infrastructure designs.

1. Arief, Mansur, Yan Akhra, Iwan Vanany. "A Robust and Efficient Optimization Model for Electric Vehicle Charging Stations in Developing Countries under Electricity Uncertainty." *Under Review*.
<https://arxiv.org/abs/2307.05470>
2. Amilia, Nissa, Zulkifli Palinrungi, Iwan Vanany, Mansur Arief. "Designing an Optimized Electric Vehicle Charging Station Infrastructure for Urban Area: A Case Study from Indonesia." *ITSC, 2022*.
<https://ieeexplore.ieee.org/abstract/document/9922278>

OPTIMIZATION UNDER UNCERTAINTY

Exploration of optimization and simulation techniques for in the context of decision-making under uncertainty.

1. Ziyad, Muhammad, Kenrick Tjandra, Mushonnifun Faiz Sugihartanto, Mansur Arief. "An Optimized and Safety-aware Maintenance Framework: A Case Study on Aircraft Engine." *ITSC, 2022*.
<https://ieeexplore.ieee.org/abstract/document/9922187>
2. Oktavian, Muhammad Rizki, Diana Febrita, Mansur Arief. "Cogeneration Power-Desalination in Small Modular Reactors (SMRs) for Load Following in Indonesia." *ICST, 2018*.
<https://ieeexplore.ieee.org/abstract/document/8528706>
3. Pujawan, Nyoman, Mansur Arief, Benny Tjahjono, and Duangpun Kritchanchai. "An Integrated Shipment Planning and Storage Capacity Decision under Uncertainty." *International Journal of Physical Distribution & Logistics Management (IJPDLM), 2015*.
<https://www.emerald.com/insight/content/doi/10.1108/IJPDLM-08-2014-0198/full/html>

Let's stay in touch

Mansur Maturidi Arief

Postdoctoral Scholar, SISL

Email: mansur.rief@stanford.edu

Web: <https://mansurarief.github.io/>

Whatsapp: [+1-734-881-0531](tel:+1-734-881-0531)