Quantitative fault tree analysis (qFTA) for autonomous systems

Edge Case Research

Making Autonomy Safer > info@ecr.ai

12th Scandinavian Conference on System & Software Safety

November 20, 2024

What do I mean by qFTA?

- qFTA: proposal for an extended use of FTA, which leverages fault trees for calculating hazard probabilities in addition to their traditional qualitative use.
- Precedent for quantitative methods in FTA

P_AND = P_A * P_B * P_C * ...

P_OR = 1 - (1 - P_A)(1 - P_B)(1 - P_C)...

Components of qFTA

- **Fault trees** corresponding to all items in the HTS
- **Hazard** Tracking System (HTS)
- Sources of **metrics data** for leaf node events

What is a hazard tracking system?

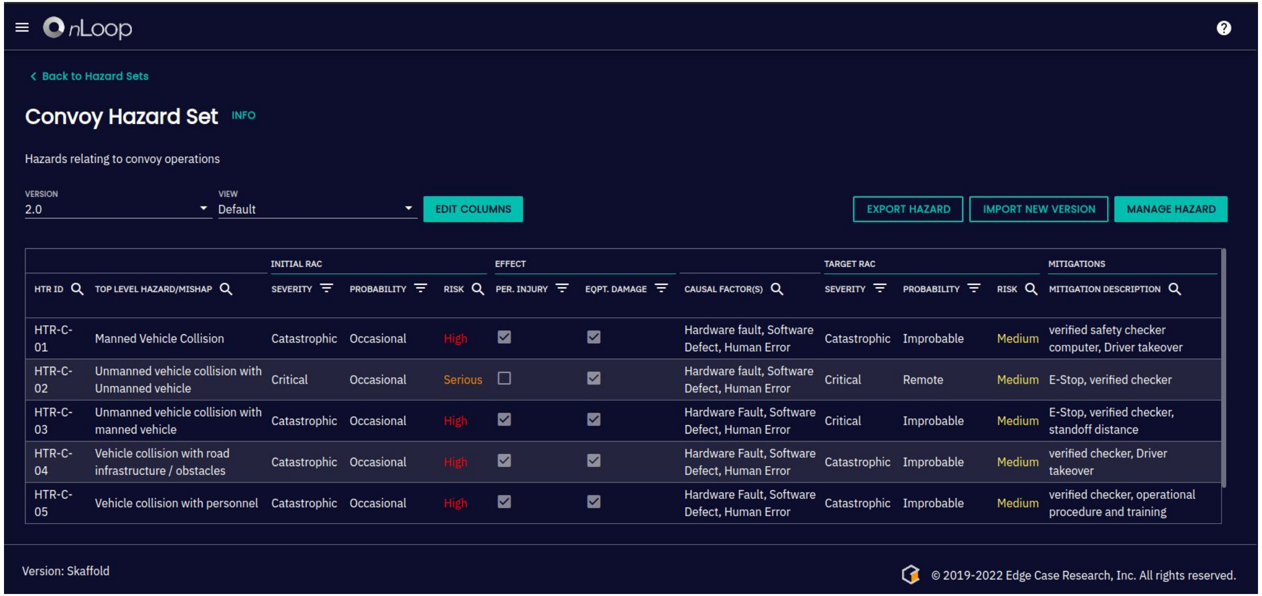
Our SMS centers on a "closed-loop" hazard tracking system (HTS) from MIL-STD-882E.

List hazards and associated mishaps.

Indicate initial, target, and current risk assessments.

Link to mitigation measures and V&V.

Document acceptance.



nLoop

Back to Hazard Sets

Convoy Hazard Set INFO

Hazards relating to convoy operations

VERSION 2.0 | VIEW Default | EDIT COLUMNS | EXPORT HAZARD | IMPORT NEW VERSION | MANAGE HAZARD

		INITIAL RAC			EFFECT			TARGET RAC			MITIGATIONS
HTR ID	TOP LEVEL HAZARD/MISHAP	SEVERITY	PROBABILITY	RISK	PER. INJURY	EQPT. DAMAGE	CAUSAL FACTOR(S)	SEVERITY	PROBABILITY	RISK	MITIGATION DESCRIPTION
HTR-C-01	Manned Vehicle Collision	Catastrophic	Occasional	High	☑	☑	Hardware fault, Software Defect, Human Error	Catastrophic	Improbable	Medium	verified safety checker computer, Driver takeover
HTR-C-02	Unmanned vehicle collision with Unmanned vehicle	Critical	Occasional	Serious	☐	☑	Hardware fault, Software Defect, Human Error	Critical	Remote	Medium	E-Stop, verified checker
HTR-C-03	Unmanned vehicle collision with manned vehicle	Catastrophic	Occasional	High	☑	☑	Hardware Fault, Software Defect, Human Error	Critical	Improbable	Medium	E-Stop, verified checker, standoff distance
HTR-C-04	Vehicle collision with road infrastructure / obstacles	Catastrophic	Occasional	High	☑	☑	Hardware Fault, Software Defect, Human Error	Catastrophic	Improbable	Medium	verified checker, Driver takeover
HTR-C-05	Vehicle collision with personnel	Catastrophic	Occasional	High	☑	☑	Hardware Fault, Software Defect, Human Error	Catastrophic	Improbable	Medium	verified checker, operational procedure and training

Version: Skaffold

Focus/Scope of qFTA

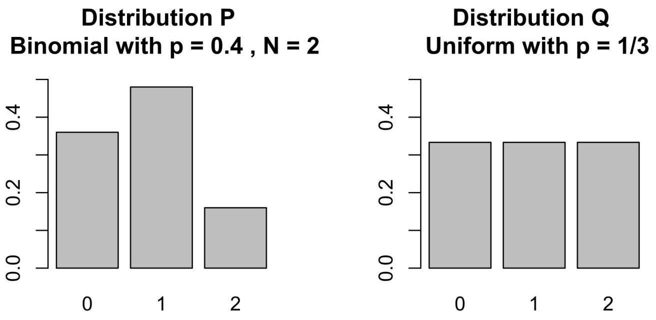
"*My risks are sufficiently low because…*":

- qFTA shows probability for all items in HTS is measure as low
- Fault trees have shown to be accurate models for failure

Focus is on system-level test in response to environmental conditions, less so on internal faults, which should be covered by other safety activities (like ISO 26262)

Using Histograms

- Single failure probabilities are fine for hardware, not for SW or ML components
- Histograms can capture more information about how components fail

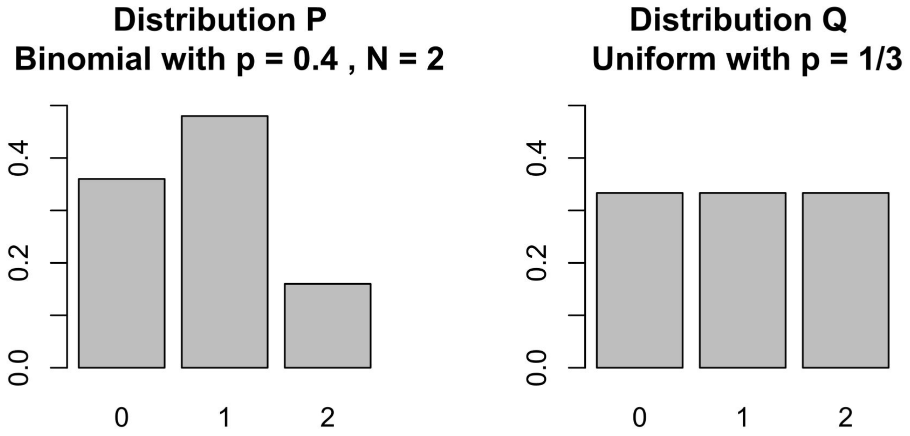


Key Concept: Kullback–Leibler divergence

Approach is based on histograms, which are compared using Kullback–Leibler (KL) divergence.

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right).$$

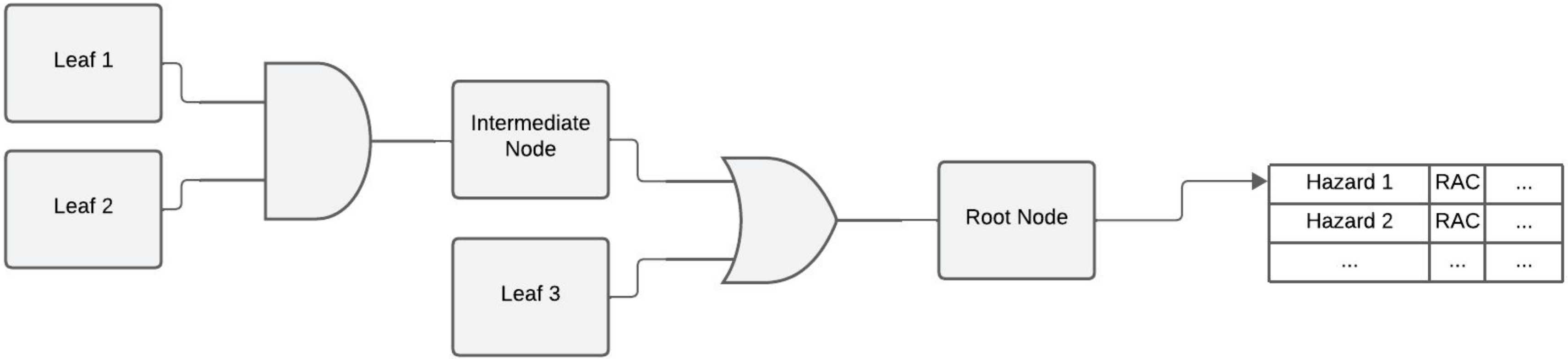
$$\begin{aligned} D_{\mathrm{KL}}(P \parallel Q) &= \sum_{x \in \mathcal{X}} P(x) \ln\left(\frac{P(x)}{Q(x)}\right) \\ &= \frac{9}{25} \ln\left(\frac{9/25}{1/3}\right) + \frac{12}{25} \ln\left(\frac{12/25}{1/3}\right) + \frac{4}{25} \ln\left(\frac{4/25}{1/3}\right) \\ &= \frac{1}{25}\left(32 \ln(2) + 55 \ln(3) - 50 \ln(5)\right) \approx 0.0852996, \end{aligned}$$



Source: https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence

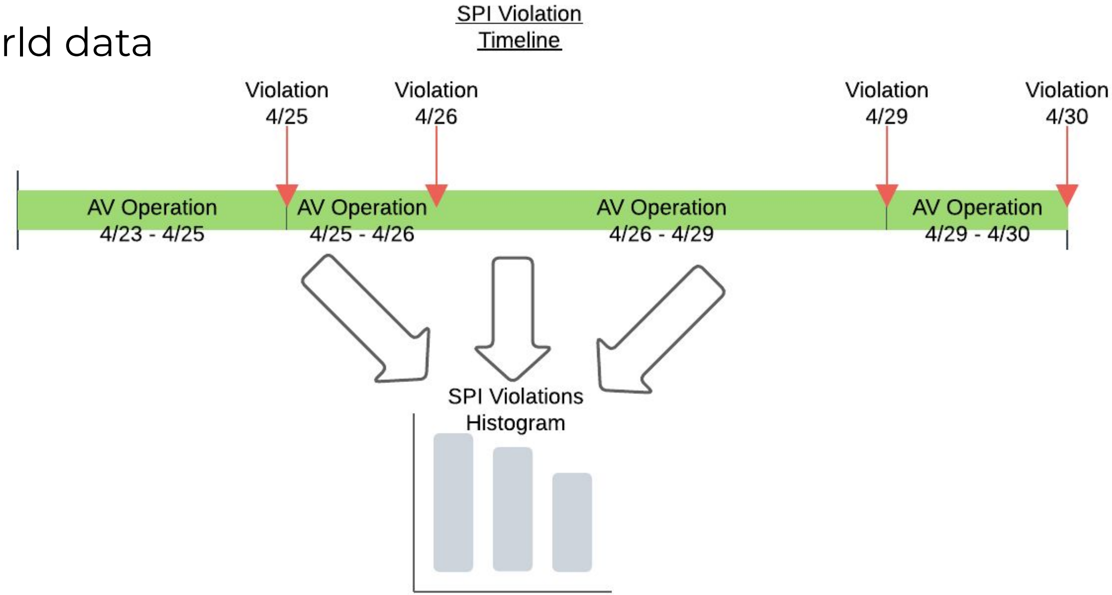
qFTA Process: Build Fault Trees

- Build fault trees based on understanding of system failures
- Trace fault trees to HTS items, which have each RAC set based on a corresponding fault tree



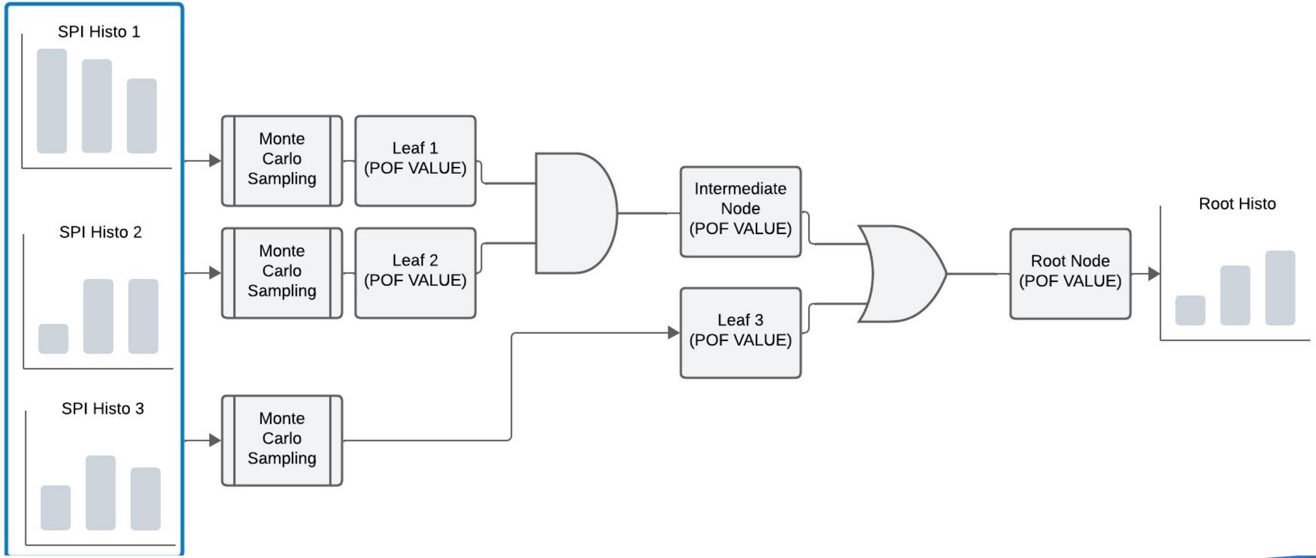
qFTA Process: Create Metric Histograms

- Gather histograms for input node failure rates
- Translation of time-series data into time-between-violations histograms
  - Simulated and real-world data



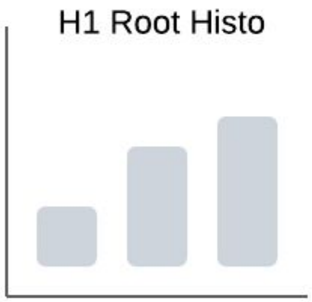
qFTA Process: Create Root Node Histograms

- Use MC sampling to calculate POF histograms for root and mid-level FT nodes
- Concrete POF values are sampled and propagated to create higher-level PDFs.



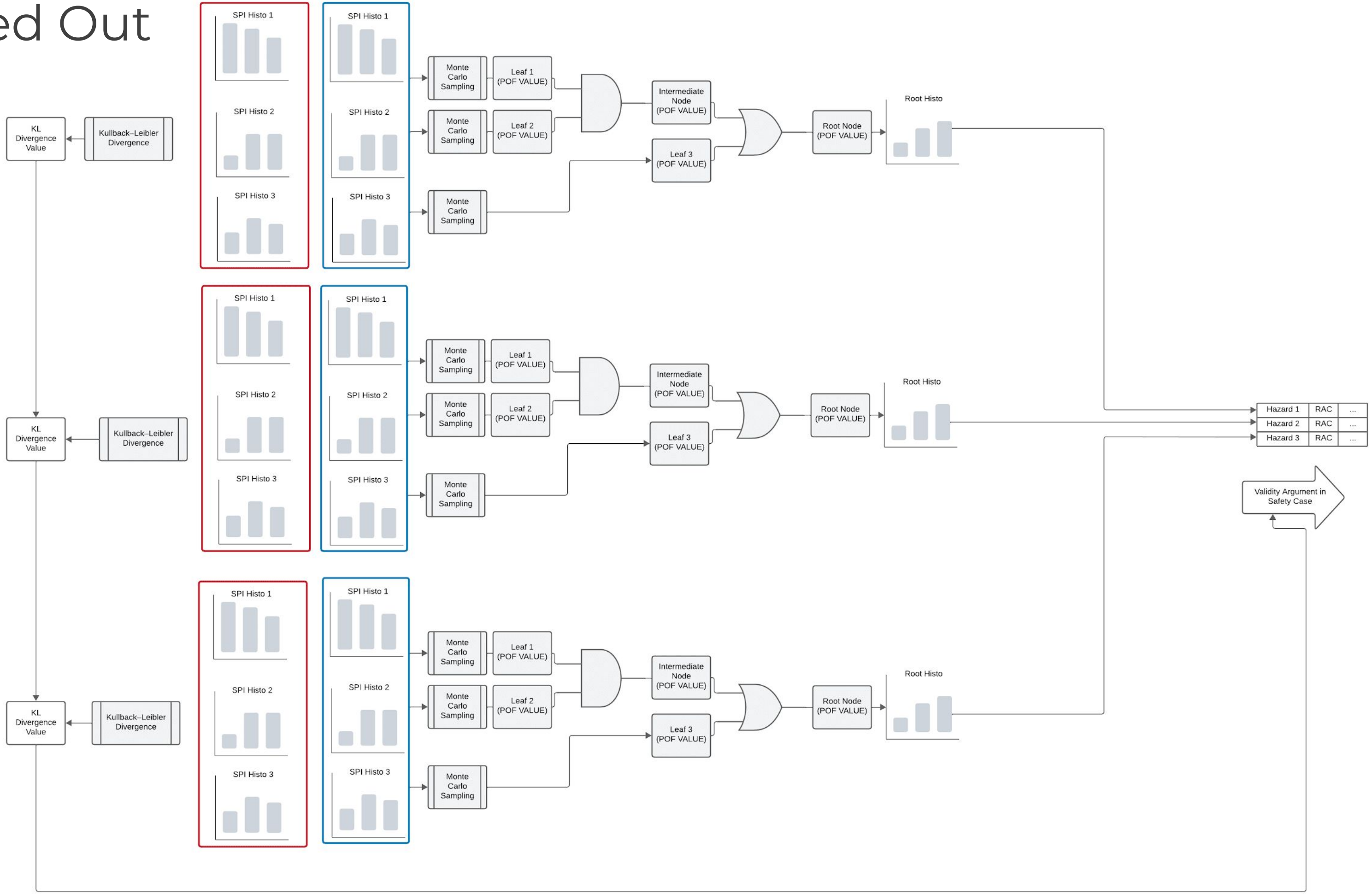
qFTA Process: Monitor RACs in HTS

- Improve system performance to reach sufficient POF rates for sign-off decision
- Acceptable expected value of root-level POFs will be determined by mission length and RAC code acceptance criteria



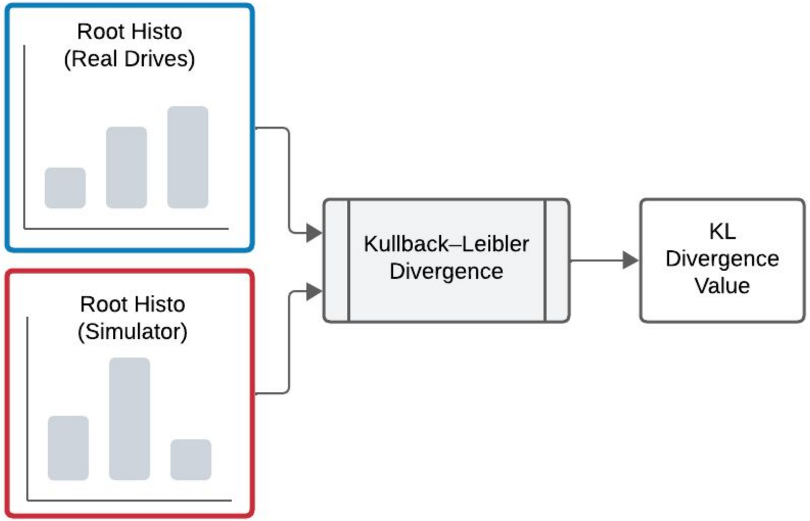
Hazard 1	RAC	...
Hazard 2	RAC	...
...	...	...

qFTA Zoomed Out



qFTA Supplemental Process: Validation of Simulator Data

- Use KL-divergence to compare higher-level POF histograms in fault trees
- Prioritize root-cause analysis based on KL-divergence value
- Reveal performance and workload gaps between simulated and real data sets



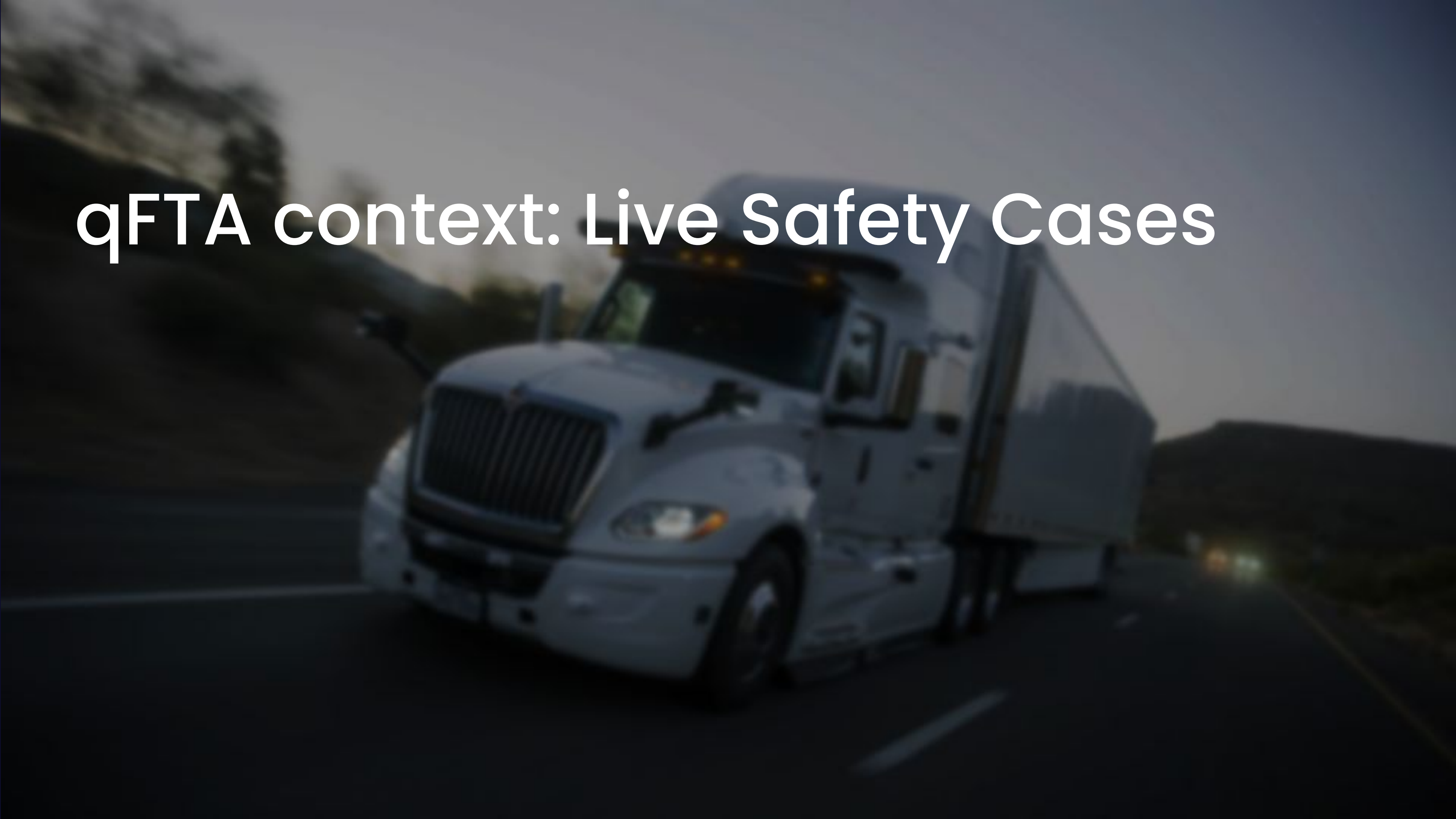
qFTA Post-Deployment

- Monitor metrics from fleet operation and cross-validate with qFTA histograms
- Use KL-divergence to compare ODD with actual OD scenario distributions
- Set thresholds on SPIs to track operational safety, anticipating unsafe behavior before loss occurs

Important Questions and Future Work

- Quality of FTs is critical for this process
- Unknown unknowns/triggering conditions may not be represented
- KL divergence quantifies distance but does not have clear path for accepting values of K as safe
- Time-between-violations is just one way of creating histograms, there might be better ones

qFTA context: Live Safety Cases



What is a safety case?

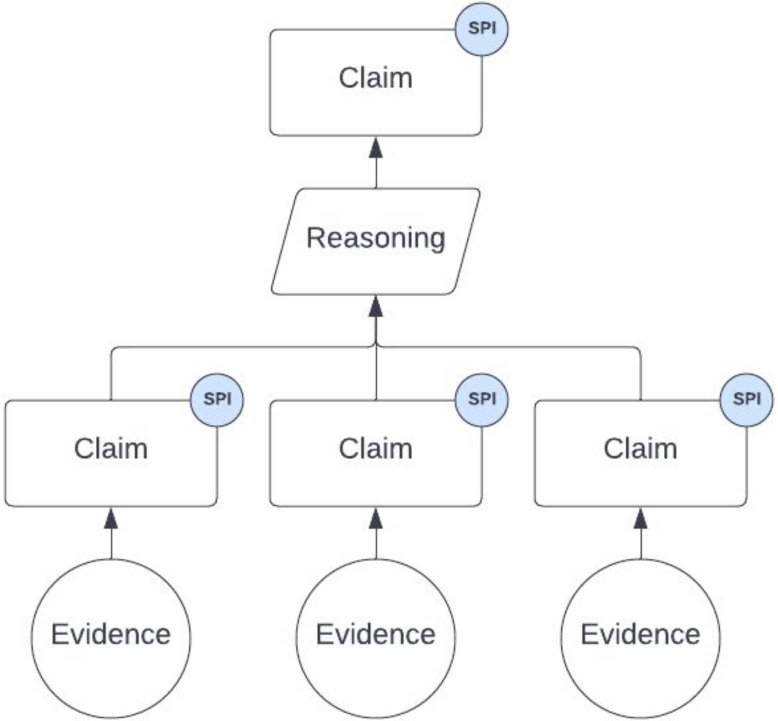
A safety case is an *argument* backed by *evidence* in support of a *claim* such as:

“Our autonomous driver is safe enough to deploy.”

Claims like these are neither formally proven nor demonstrated by testing alone.

We instead rely on arguments about process rigor, standards conformance, analyses performed, and statistical risk assessments.

Safety-critical industries such as nuclear, rail, and aviation use safety cases to make risk-acceptance decisions.



What are safety performance indicators?

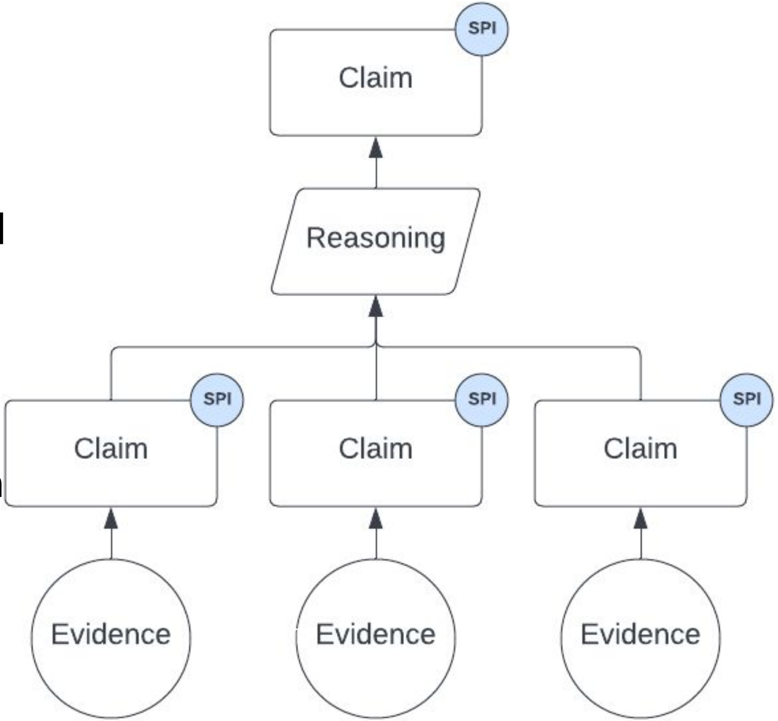
Safety performance indicators (SPIs) are metrics with thresholds that trace to claims. They're used to detect edge cases and risks to the safety case.

When a SPI is violated (i.e., the metric crosses its threshold) its linked claim is potentially invalid.

Technical SPIs trace to autonomy functions such as perception, prediction, planning, and control.

Operational SPIs trace to safety management systems, processes, and safety culture.

Safety case + SPI + data = **live safety case**



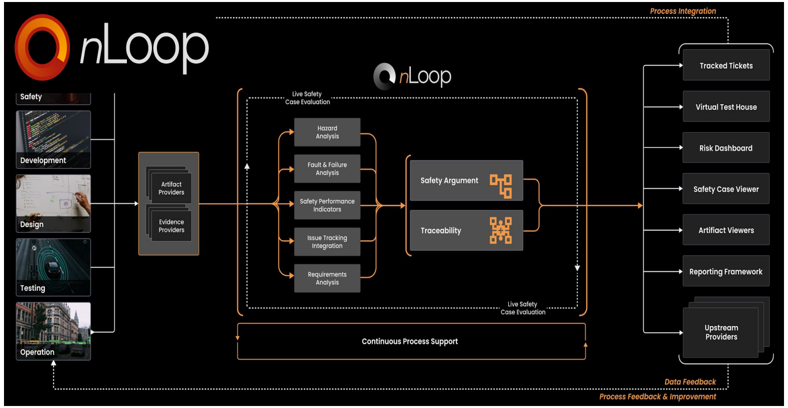
Why is a "live" safety case important?

Safety cases rely on inductive reasoning.

Novel systems and unrestricted environments lead to uncertainty in any safety case.

Need to review feedback constantly, both during development and in operation.

SMS defines and enforces how this review takes place, but analysis tools are required.



nLoop Demo

Thank You

Edge Case Research

Noah Carlson

ncarlson@ecr.ai

Applications of qFTA

- Challenge of ***quantification of release criteria*** for autonomous vehicles
- Need for ***simulator-based data*** to accrue necessary mileage
- Challenge of post-deployment monitoring that contains ***leading indicators*** of safety

Benefits of the Approach

1. Fault trees provide leading indicators of safety
2. Intermediate event rate calculation can provide validation evidence without need of metrics for rarer, unsafe, top-level events
3. Allows for simulated data in safety-specific context, in relation to hazards
4. Histogram sampling avoids need for concrete failure rates, which may not be available (or knowable) for ML components
5. SPI histograms can be gathered for sign-off decision, and can be used for continuous validation post-deployment