

Waymo's Safety Case

Trent Victor, PhD
Director of Safety Research
& Best Practices



Safety Case Structure

Safety Case Structure

GOAL
(Overarching statement)

The top-level goal of Absence of Unreasonable Risk. **Safety** is defined in ISO as **Absence of Unreasonable Risk (AUR)**

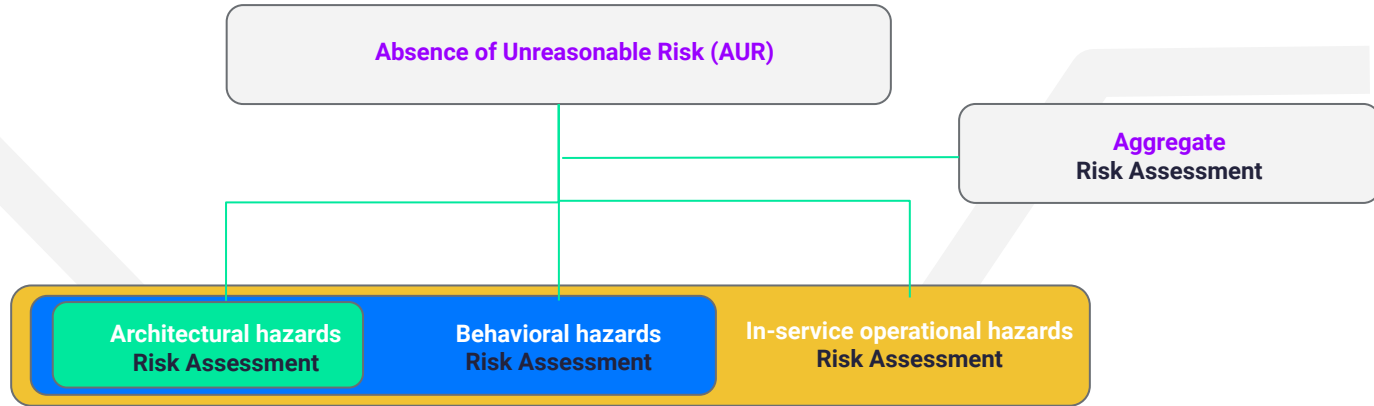
LOGICAL ARGUMENT
(Decomposing the statement)

Building a Credible Case for Safety: Waymo's Approach for the Determination of Absence of Unreasonable Risk (March 2023)

EVIDENCE
(Compelling proof)

Waymo's Safety Methodologies and Safety Readiness Determinations (October 2020)

Decomposing AUR

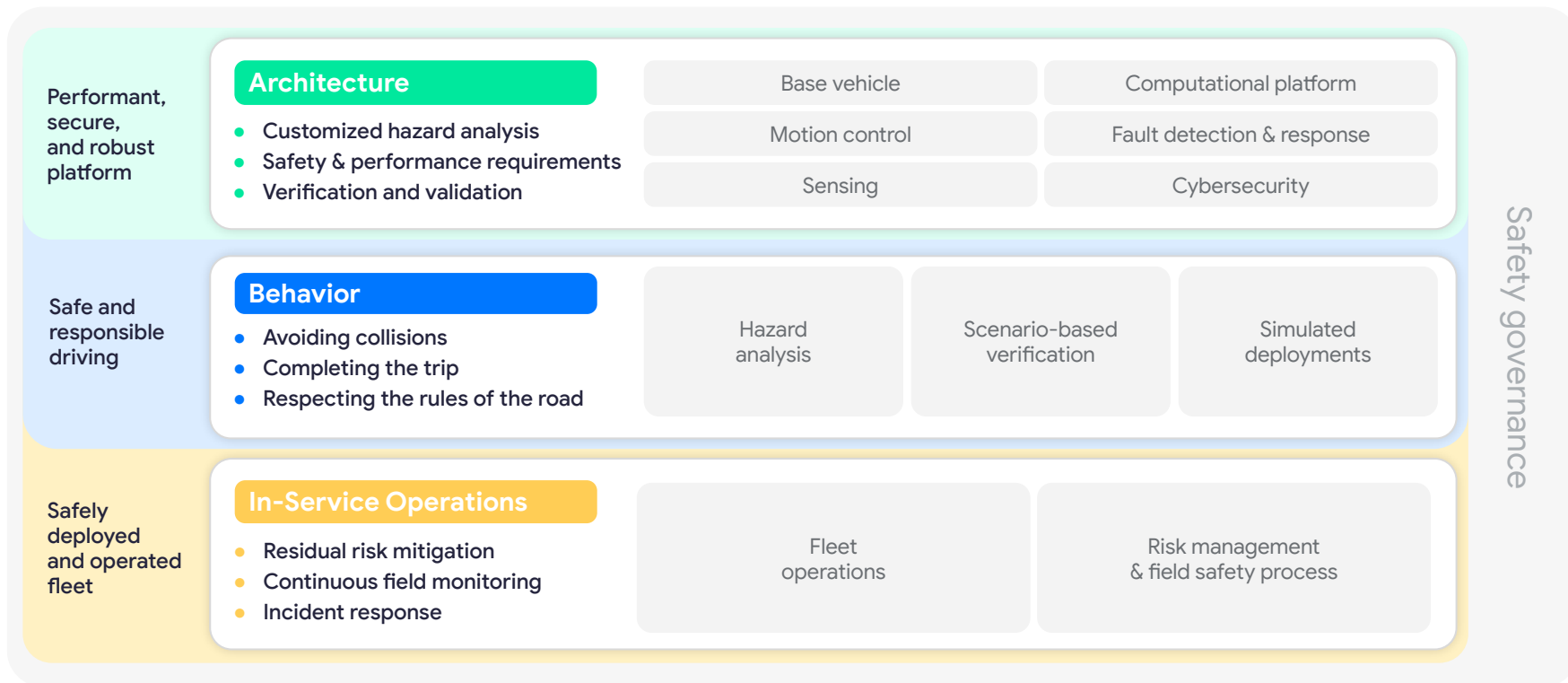


Architectural hazards: those associated with potential sources of harm inherently embedded within the platform because of architectural choices.

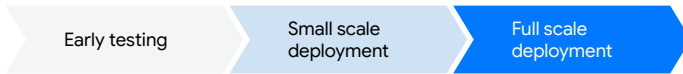
Behavioral hazards: those associated with potential sources of harm resulting from the ADS's displayed driving behavior, whether intended or unintended.

In-service operational hazards: those associated with potential sources of harm resulting from the fact that the ADS operates in a complex ecosystem, and that do not belong to the other two categories.

A layered approach to safety

waymo.com/safety


Determination of absence of unreasonable risk

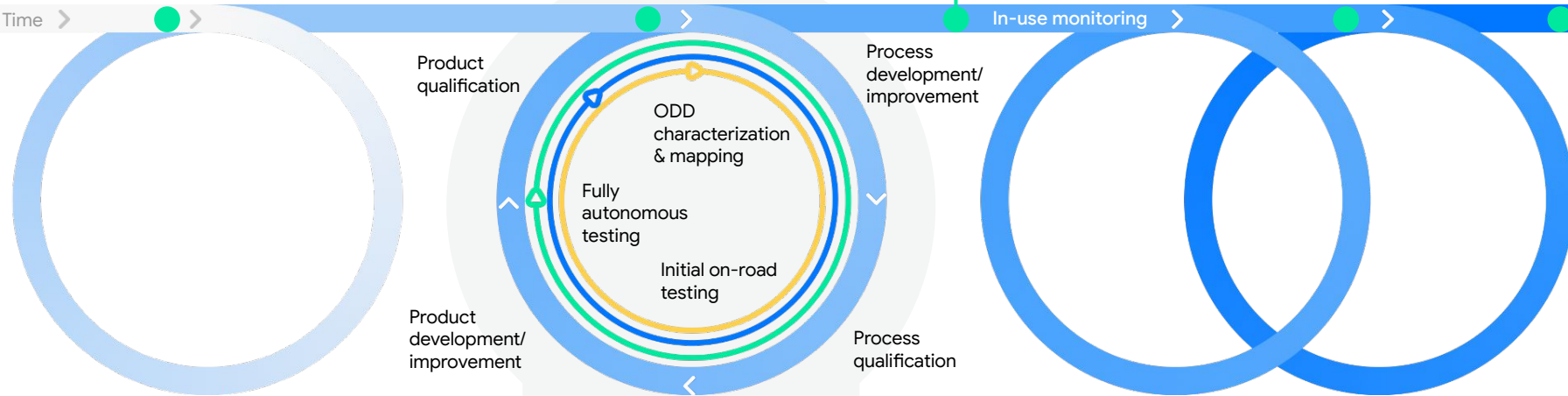


- Architecture
- Behavior
- In-service Operations

Safety as an acceptable prediction and/or observation

Deployment Readiness Review

Time >



Process and Product Continuous Refinement

Context Definition:

- Vehicle configuration (platform)
- Operational configuration
- ODD selection
- Sought deployment scale

Safety as an emergent development property

Safety as continuous confidence growth

Key Evidence

Swiss Re led study found Waymo is significantly safer towards other road users. In over 3.8 million miles with no human in the driver seat:

- Bodily injury claim frequency reduced 100%
- Property damage claim frequency reduced 76%

Di Lillo, et al., 2023

Either inherently avoided reconstructed fatalities, or avoided/mitigated them with collision avoidance

Scanlon, et al. 2021; Kusano, et al., 2022

In our first million rider-only miles:

- No reported injuries
- No collisions with pedestrians or cyclists
- Encountered dangerous human driving

Victor, et al., 2023



Better than a non-impaired, attentive human driver

Scanlon et al, 2022; Engstrom et al, 2022; Kusano, et al., 2022.

Waymo follows speed limits; humans speed 27-47% of the time.

Waymo blog

Safety Case approach & toolkit

Favaro, et al., 2023

Holistic safety readiness methodologies





Webb, et al., 2020

Balancing Aggregate- and Event-level Reasoning

The usage of safety performance outcomes for the determination of safety can lead to over-indexing on aggregate performance indicators that inadvertently conceal the presence of undesirable levels of risk in individual events or scenarios.

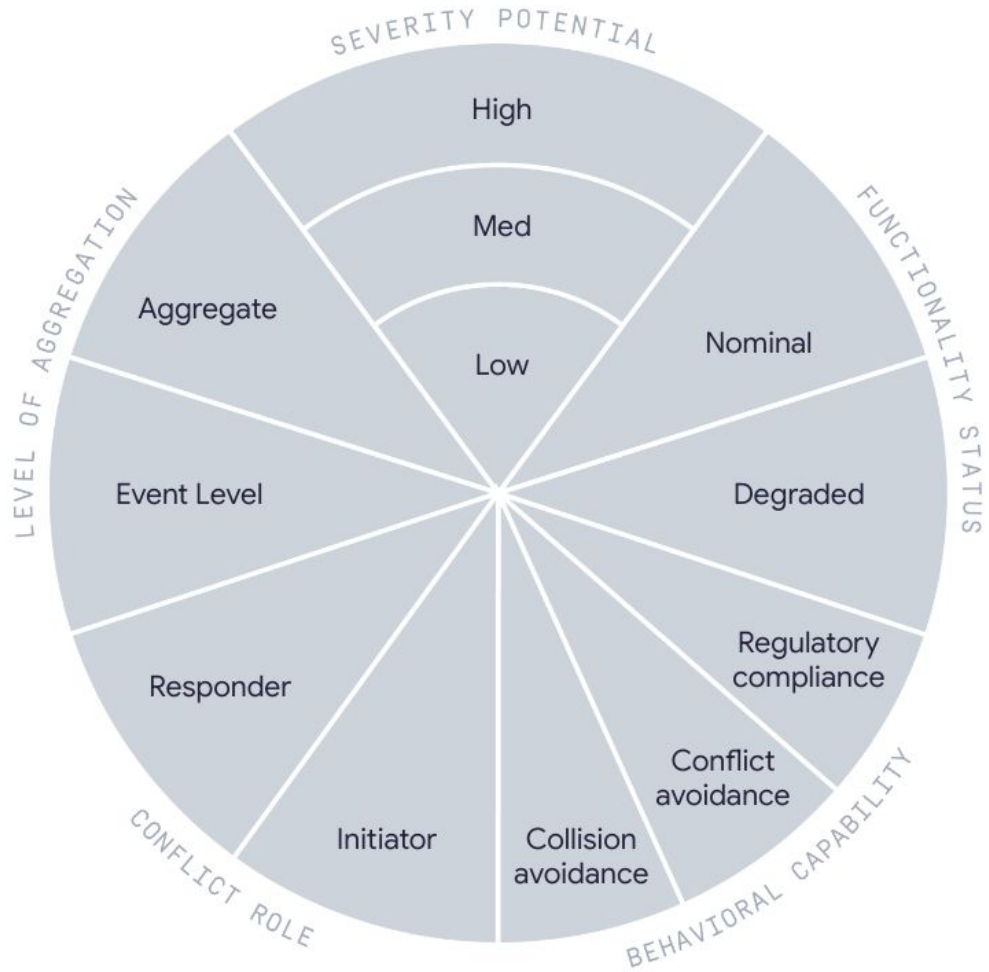
The assertion that the Waymo Driver is successful at reducing injuries and fatalities is thus grounded in analyses that go beyond the prediction of fatality rates.

Acceptance Criteria Enabling Event-Level Reasoning

		Included	Not Included
Acceptance Criteria Enabling Aggregate-Level Reasoning	Included	 <p>An appropriate balance coming from the inclusion of both event level and aggregate level indicators helps ensure that the risks for a given scenario category are being captured. It also enables the evaluation of single undesirable behaviors that a developer needs to consider to show that residual risk is as low as reasonably possible.</p>	 <p>An argumentation based only on aggregate criteria may not capture some risk posed by the ADS in individual scenarios/situations. Furthermore, confidence in aggregate rates pre-deployment is constrained by the available data collected during testing.</p>
	Not Included	 <p>There are infinitely many operational scenarios that an ADS will be exposed to. Establishing a safety argument only on event-level instances precludes the holistic assessment of residual risk. Furthermore, aggregate-level criteria can provide validation for those trends observed from event-level indicators.</p>	 <p>No argumentation possible in the absence of acceptance criteria, since Absence of Unreasonable Risk is a necessary goal for ADS deployment</p>

Acceptance Criteria Framework for AUR Behavioral Evaluation

We can define the minimum set of dimensions of interest to state completeness of the set of acceptance criteria and establish credibility



Conclusions

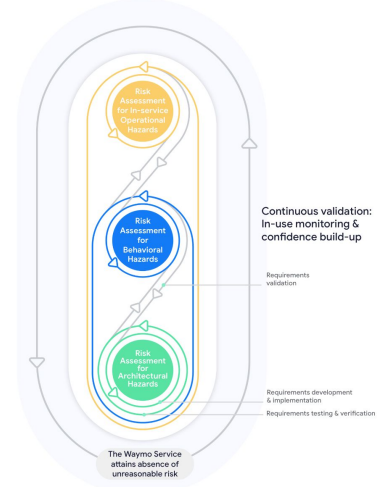
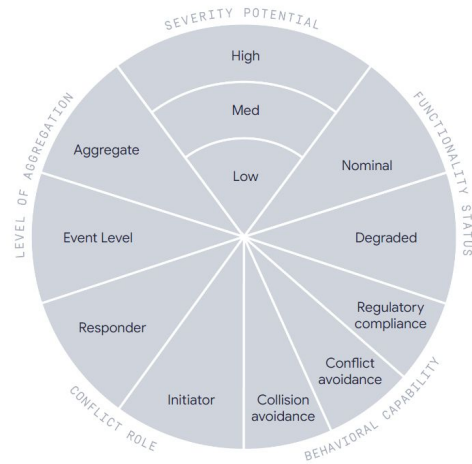
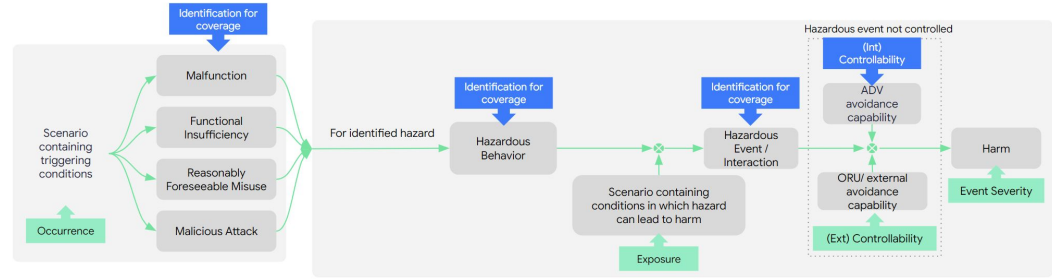
The determination of **safety** is a **risk assessment** process.

The **layered approach to safety** consist of decomposing the determination of absence of unreasonable **risk** into **architectural, behavioral, and in-service operational hazards**.

Each of these hazard categories requires a set of **explicit risk acceptance criteria**. Setting appropriate criteria relies on:

- A sufficiently exhaustive list of hazards,
- Appropriate performance indicators
- Acceptance criteria framework with appropriate dimensions of interest

Safety = Absence of Unreasonable Risk



Closing / Thank you

Trent Victor, PhD
Director of Safety Research
& Best Practices

