

Independent Verification for AV Performance Metrics

How motion-based validation strengthens existing AV metrics and the evidence behind them

Executive Summary

AV developers already measure safety, reliability and performance. The question is how much confidence to place in a result that depends on the same sensors, software, logs and reporting path as the system being evaluated.

BoundaryScore™ adds an independent physical reference. Observable motion is checked against what actually happened and reconciled with machine-reported state. Agreement validates the physical component and increases confidence in the associated machine-state record, while the underlying evidence is preserved for traceability and audit. When the applicable requirements are met, the result can be designated BoundaryScore™ Verified.

IN THIS PAPER

1. Existing metrics

Why independent validation matters.

2. Verification layer

What motion can independently support.

3. Metric families

Where validation contributes.

4. Evidence levels

How claims are bounded.

5. Evidence chain

What makes results traceable and auditable.

6. BoundaryScore™ Verified

How verified evidence supports internal and external confidence.

1. The market already has metrics

Different organizations measure different things for good reasons. Waymo reports rider-only exposure and crash outcomes against human benchmarks. Tesla reports fleet mileage and collision measures for FSD (Supervised). NHTSA collects ADS/ADAS crash reports but notes that its summary data are not normalized by exposure, vehicle counts or operating domain. IIHS uses controlled tests, real-world insurance data and safeguard evaluations, while SAE J3016 defines levels of driving automation by task and fallback responsibility.[1-5]

Those measures answer different questions. The practical issue arises when a reported result cannot be checked against an independent reference or when an outside decision-maker cannot determine how the result was produced.

2. BoundaryScore™ as the independent verification layer

BoundaryScore™ begins with the metric or performance claim already being used. For observable motion, an independent physical reference can validate position, velocity, acceleration, trajectory, braking, stopping behavior and timing, then reconcile those observations with machine-reported state.

Other inputs - such as operating mode, intervention state, fault state, classifications or selected ODD attributes - may still come from the participant or another external source. BoundaryScore™ keeps those boundaries explicit so independently observed facts are not confused with reported inputs or unsupported conclusions.



3. Metric families and evidence support

The metric families below show where independent physical evidence can strengthen existing AV performance measures without replacing the definitions already in use. BoundaryScore™ can directly validate observable motion and timing, while participant-supplied or external inputs remain explicitly identified when they are necessary to interpret a metric. This separation makes clear which parts of a result are independently supported and which still depend on system state, classification or contextual records.

Metric family and typical measures	What independent evidence can support
Exposure — miles, hours, trips, operating time	Independent motion/time can validate distance and duration; operating mode may require participant input.
Collisions — events, rates, severity proxies	Motion can corroborate event timing and dynamics; injury, fault and classification may require external/contextual records.
Interventions — disengagements, takeovers, overrides	Motion can corroborate timing and vehicle response; intervention state generally comes from participant data.
Braking — deceleration, stopping behavior, AEB response	Independent motion can validate speed change, deceleration, stopping distance and timing.
Trajectory — path, lane/route adherence, stability	Independent trajectory can be compared directly with reported path or a defined corridor.
ODD performance — road, weather, geography, speed	Motion and exposure can be validated; ODD labels and some context require participant or external data.
Fallback behavior — minimum-risk maneuver, stop, handoff	Independent motion can validate the executed maneuver; trigger and system state may require participant input.
Reliability — availability, interruption, fault-free operation	Observed continuity can contribute; internal fault and availability definitions usually require participant data.
Mission completion — trip/task or route completion	Observed movement can support completion evidence; task success may require application-specific inputs.

4. Verification must be explicit about its limits

A credible verification framework should not treat every metric as equally observable. The output should show the level and source of support so a directly measured physical quantity is not confused with a result that depends on participant-supplied denominators, classifications or internal system state.

Verification level	How it should be represented
Directly observable	Independent motion establishes the physical quantity or behavior. State the verified component directly, retaining uncertainty, time and context.
Partly participant-supplied	Independent evidence supports part of the result while another input comes from the participant. Identify both explicitly.
Event corroboration	An independently observed event can be reconciled with a reported event or sequence. Do not infer cause, fault or safety unless separately supported.
Context-dependent	Interpretation depends on ODD, weather, roadway, nearby actors or other context. Preserve source and context.
Not independently observable	The claim depends on proprietary state or information unavailable to independent measurement. Record only what the evidence supports.

5. The evidence chain is part of the product

Independent verification is more than a second measurement. The record should preserve the evidence required to understand, reproduce and audit the result:

- Definition - exactly what is being measured or claimed, including numerator and denominator where applicable.
- Evidence - independent motion observations, machine logs, participant inputs and external sources used in the result.
- Timing and integrity - timestamps, authentication, security controls and other mechanisms that preserve when evidence was generated and how it entered the record.
- Context - operating conditions, exposure, ODD information and relevant exclusions.
- Methodology - reconciliation logic, thresholds, uncertainty treatment and calculation rules.
- Provenance and version - where each input originated, how it was transformed, and the method, metric definition and system version to which the result applies.

Because the physical reference is generated independently from the system being evaluated, the resulting evidence chain is harder to alter or reconstruct after the fact. That makes the metric easier to trace, reproduce, audit and defend.

6. BoundaryScore™ Verified

BoundaryScore™ Verified is the designation applied when a defined metric or performance claim meets the applicable verification requirements. Independent motion validates the physical behavior directly. When that motion agrees with the machine-reported motion state, the match also provides stronger evidence that the associated machine-state record is trustworthy. The verification record preserves what was observed, what was reported, the conditions of the result and the evidence supporting the conclusion.

Greater confidence inside the company

Engineering, safety, quality and executive teams can use the same verification record to understand where a metric is independently supported, where it depends on reported inputs and where uncertainty remains. That gives internal reviewers a stronger basis for trusting, challenging and repeating the result.

Greater credibility outside the company

The same record can support customers, insurers, regulators, partners, standards bodies and other authorized stakeholders that need evidence they did not generate themselves. They can see what was independently observed, what was supplied by the participant, the conditions under which the conclusion is valid and the evidence trail supporting it.

The core proposition

Validate existing metrics against independent motion. Preserve a traceable evidence chain. Increase confidence internally and credibility externally.

Sources

- [1] Waymo, Safety Impact, data and methodology current through March 2026.
- [2] Tesla, Full Self-Driving (Supervised) Vehicle Safety Report and methodology.
- [3] NHTSA, Standing General Order on Crash Reporting; agency notes that summary incident data are not normalized by exposure.
- [4] Insurance Institute for Highway Safety, Advanced Driver Assistance and vehicle-rating methodologies.
- [5] SAE International, J3016_202104, Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.