

When the sources are judgements, not numbers

Several models disagree about what they are looking at. Which one do you act on?

AT A GLANCE

The problem. Combining verdicts is not like combining measurements. Shared failure modes can make agreement less informative than it appears.

What this does. Measures how far a source sits from the others rather than from its own history, and separates reaching a conclusion from being permitted to act on it.

Who it is for. Anyone running several models or human assessors over the same input and having to act on the result — autonomy stacks, security operations, screening and triage, multi-agent AI systems.

Maturity. Implemented in a company-funded reference implementation and exercised against constructed scenarios. **Not yet proven in the field.**

Status. Covered by US provisional patent application **64/161,448**, filed 24 September 2026. The fourth of four applications filed 2026, carrying the others' discipline into sources that emit judgements rather than measurements; the earlier three cover disagreeing sources of continuous data, a declared minimum of participating sources, and accounting for what was received and what was not. Patent pending; none examined or granted.

What combining verdicts assumes

A dozen sensors reporting range can be combined by arithmetic: the answer sits between them, and a drifting source pulls the result by a calculable amount. Judgements are different, though not because they cannot be combined. Averaging the *labels* is meaningless — there is no midpoint between *benign* and *hostile* — but averaging the class probabilities behind them, usually called soft voting, is a well-established ensemble technique and frequently the right thing to do. Majority vote, stacking and calibration all have their place. The difficulty is not the arithmetic but the condition each method needs.

Shared failure modes can make agreement less informative than it appears. Four models against one looks decisive, and it is — if the four arrived at their answers by different routes. It is also what a single correct source looks like when the other four share a training set, a sensor modality, or a bad afternoon of weather. Some methods address this directly: stacking learns the correlation structure from held-out data, correlation-aware fusion models it explicitly. Both depend on that structure having been characterised in advance and still holding on the day.

Confidence is easy to take at face value. A model that is confidently wrong reports high confidence, and soft voting weights it accordingly. Calibration helps and does not settle it: a source calibrated on data that looks nothing like today is calibrated for the wrong day.

Underneath both is one habit worth naming. **A source's reliability is frequently read from the source** — its own probability, its own history, its own self-report. How it stands against the others does not. That is not the only evidence available, and it survives a source being wrong about itself.

What this capability does

Disagreement is measured against a declared scale. The verdicts a system can emit are enumerated in advance and the distance between any two is declared rather than assumed, so "how far apart are these judgements" is a quantity, not a show of hands.

A source is compared against the consensus formed without it, so it cannot help set the standard it is then measured against. This matters more the fewer sources there are.

Rate of change is bounded by declaration. For each channel, an operator declares how fast it may move across the verdict scale, from what is known about the thing being observed. A change beyond that bound is recorded as **exceeding the declared bound** and is not treated as news.

That bound is a configured assumption and it can be wrong. A real event can produce a genuinely abrupt change, and where the bound is set too tight a true detection will be recorded as exceeding it. Exceeding the bound is never recorded as a proven fault in the channel — only as what it is, a reading outside a declared assumption. The bound is stored with the decision.

Concluding and acting are separated. Where an action commits something that cannot be undone, its release is held behind a gate of its own, so a marginal consensus can produce a recorded conclusion and no irreversible act.

The record is hash-chained. What each source said, which were counted, what was resolved, what was authorised or withheld — a later edit breaks the chain, found by recomputing it against a separately held anchor. Editing is detectable, not impossible, and an anchor the record's host can rewrite proves correspondingly less.

An illustrative scenario

This is not a customer story. No deployment described here has taken place. The behaviour below follows from how this configuration was set up; a different one behaves differently.

Four detection models watch the same feed. Three report the scene as routine. The fourth reports a hazard, and has done so for six consecutive intervals. It is compared against the consensus of the other three: the distance is large, and it has persisted rather than flickered.

Persistence does not mean the dissenter is right. A source that is persistently wrong — miscalibrated, pointed at a reflection, running on stale weights — produces the same pattern. Nothing here tells those apart, and the method does not try to.

What it does is narrower. It establishes that the four sources are not in a state this configuration accepts as supporting the action, records that with the distances that produced it, and holds the action that would otherwise have followed **under the gate this deployment**

declared. The state carries a declared alert action rather than the irreversible one, and a person's acknowledgement is bound into the next record.

Getting the interval in front of the right person is the deployment's job. The reference implementation declares the alert action and records the acknowledgement; it ships no review console, and any queue, roster or escalation path around it is a proposed workflow, not something demonstrated.

Had the fourth model instead jumped from *clear* to *hazard* in one interval, that change would have exceeded the bound declared for that channel and be recorded as exceeding it — not as a detection, and not as a fault.

What the record supports. Each record carries a digest of the configuration it ran under, the declared bound included, so a run under a different configuration cannot pass as the original. Re-running an interval against a different bound to see what would have happened is a counterfactual — not evidence of what happened, and a proposed review workflow rather than one the reference implementation performs.

What this is not

It is not a detector. It does not find hazards, classify images or replace the models it sits over. It decides what to do when they disagree.

It does not establish truth. It cannot identify which source is correct, and disagreement is not evidence that any source is right.

It does not make correlated sources independent. If every source shares a blind spot there is no disagreement to measure. Independence is an assumption this depends on, not one it supplies or verifies.

It does not replace ensemble methods. Soft voting, calibration and stacking remain useful, and this can sit over them.

It is not tamper-proof, and not a safety certification. Alteration is detectable, not prevented, and nothing here has been assessed against DO-178C, ISO 26262, IEC 61508 or any comparable standard.

It is not deployed, and it is not a granted patent. One provisional application covers this work; a provisional is not examined and confers no enforceable right by itself.

If several models disagree over the same input and a vote nobody is comfortable with currently settles it, that is the conversation we want. Bring a nonconfidential example, and we will say whether this would earn its place — including when it would not.

info@rvtechconsulting.com · RV Technology Consulting, Las Vegas, Nevada

Covered by US provisional patent application 64/161,448, filed 24 September 2026. Four US provisional patent applications filed 2026. Patent pending; none examined or granted. The scenario above is illustrative and describes no actual deployment. This document describes capability only and is not a technical specification.