Confidence-Governed Actuation: Graduated Modes for Physical Systems

by Nick Clark | Published April 25, 2026

Binary Permit/Suppress Has Reached Its Architectural Limit

Functional safety architectures developed for industrial control and automotive ECUs are built on a binary execution model: a command either passes safety checks and is executed, or it fails and is suppressed. ISO 26262, IEC 61508, ISO 13849, and similar standards encode this binary into safety integrity levels (SIL/ASIL) that gate actuator authority.

Binary permit/suppress works for systems where the choice is between 'do this' and 'do nothing.' It does not capture the actual decision space of an autonomous physical system, which routinely faces choices between full execution, advisory display only, partial commitment, deferred action, harm-minimizing deviation, and emergency override of normal authority.

When the choice is forced into binary form, the architecture suppresses too aggressively (legitimate actuations get refused at the boundary) or too permissively (unsafe actuations get approved because the alternative would suppress a needed action). Both failure modes are observed in current L4 fleet telemetry.

1. The Primitive: Eleven Graduated Modes

Confidence-governed actuation produces a graduated mode set evaluated for every actuation request: simulated, advisory, consultative, shadowed, partial, constrained, stage-gated, deferred, full, emergency-accelerated, and emergency-overridden. The mode is selected by a composite admissibility evaluator that combines authority, evidential weighting, capability envelope, temporal scope, and disposition into a single deterministic computation.

Each mode is a structurally distinct outcome. Advisory mode displays the contemplated action to a human operator without commanding the actuator. Shadowed mode logs the contemplated command and runs verification in parallel with continued human or fallback control. Partial mode commits a fraction of the requested authority. Stage-gated mode commits in successive bounded stages with intermediate evaluations.

Mode selection is not a heuristic. It is a deterministic function of the admissibility computation against the governance policy. The same input produces the same output. Every selection is recorded in lineage with the supporting computation.

2. Preemption Budget: Rate-Limited Override Authority

Emergency preemption authority — the ability to override normal actuation gating — is itself rate-limited under a budget. A unit may invoke preemption no more than N times within a governance-policy-defined window, and each invocation expires after a bounded duration unless explicitly renewed.

This solves the recurring problem in safety-critical systems that emergency overrides become routine, eroding the structural meaning of 'emergency.' Under a budget, preemption invocations consume a finite resource, and excessive consumption raises governance-flagged events that propagate through the mesh.

Preemption budget is configurable per authority class: a regulatory authority may grant a fleet a higher budget than a peer authority. Budget consumption is recorded with the originating authority's signature, producing tamper-evident audit of every override.

3. Reversibility-Aware Staged Commitment

Actuator commands differ in reversibility. Steering input is highly reversible. Brake application is reversible up to a point. Airbag deployment is irreversible. The execution primitive evaluates reversibility for every contemplated commitment and prefers reversible paths when feasible.

Stage-gated mode commits irreversible authority in successive bounded stages. A landing aircraft can commit to descent (reversible), then to flare (less reversible), then to wheel touchdown (committed). At each stage, the admissibility evaluator runs again with updated environmental observations. The stage progression is structurally distinct from a continuous control output.

Reversibility is computed against a governance-policy-defined classification per actuator type. The classification is published, auditable, and configurable; new actuator types receive classifications through governance-credentialed updates rather than firmware changes.

4. Governance-Policy-Configurable Harm Ordering

When all available actuations produce some harm, the primitive selects the actuation that minimizes harm under a governance-policy-configurable entity-class harm ordering. The ordering specifies relative weighting between protected entity classes (pedestrians, cyclists, occupants, property, the unit itself) and is signed by the governing jurisdiction.

This solves the trolley-problem liability gap that has dogged autonomous-vehicle ethics for a decade. Current AV stacks either hardcode a harm ordering (which transfers the ethical authority to the manufacturer) or refuse to articulate one (which leaves liability unallocated). The governed primitive externalizes the ordering: state DOTs and insurers configure it, the unit executes it, and the lineage records every harm-minimization deviation with the policy under which it was evaluated.

The harm-ordering mechanism is also extensible to non-vehicular contexts: industrial robotics with multiple object classes, medical autonomy with patient-vs-staff prioritization, defense systems with combatant-vs-noncombatant classifications. The mechanism is the same; the configuration changes.

5. Post-Actuation Verification With Mesh Broadcast

After an actuator command is committed, a verification stage observes the effect (sensor readback, environmental response, downstream telemetry) and compares it to the predicted effect. A discrepancy classifier produces one of: nominal, expected-noise, anomaly, fault, or adversarial-interference. The classification is recorded in lineage and broadcast through the governed mesh.

Mesh broadcast of actuation state is a structural change from current architectures, where actuation is a private internal state of the executing unit. Broadcasting actuation state allows other operating units, infrastructure agents, and regulatory authorities to observe what is being done in their environment, supporting cross-unit coordination, intervention, and audit.

Broadcast does not require human-readable intent disclosure. The broadcast is structural: which actuator, what magnitude, what authority gated it, what mode was selected. Privacy and competitive concerns are addressed through governance-policy-configurable redaction rules rather than by withholding the broadcast entirely.

6. Composition Across Coupled Actuators

Real autonomous systems coordinate multiple actuators (steering, brake, throttle, signaling, attention) under a single governance frame. The primitive composes across coupled actuators: a single admissibility evaluation can produce a vector of mode selections, with cross-actuator constraints (e.g., 'no brake commitment without simultaneous signaling') enforced structurally.

Cross-actuator composition is also governance-configurable. A regulatory authority can mandate that any lane change actuation include simultaneous turn-signal actuation; a port authority can mandate that any container handling actuation include simultaneous custody-transfer broadcast. The cross-actuator constraint is signed and propagated through the mesh.

7. What This Is Not

This is not ISO 26262. ISO 26262 specifies safety integrity levels and process requirements for binary safe-or-unsafe gating. The governed primitive consumes ISO 26262-classified actuators but produces graduated modes the standard does not specify.

This is not Mobileye RSS. RSS encodes formal safety distance constraints. The governed primitive can integrate RSS as one of its admissibility factors but produces mode-graduated output where RSS produces binary 'safe' classifications.

This is not Model Predictive Control. MPC computes optimal actuator trajectories. The governed primitive evaluates whether and how to commit to whatever trajectory is requested by upstream planning. The two compose: MPC produces a trajectory request; the governed primitive selects the mode under which to commit to it.

Conclusion

Confidence-governed actuation reframes execution from binary permit/suppress to graduated multi-mode commitment under deterministic admissibility. The preemption budget makes override authority rate-limited and auditable. Reversibility-aware staged commitment converts irreversible actions into bounded sequences of reversible decisions. Configurable harm ordering externalizes the ethical authority that current architectures hardcode or omit.

This is the keystone execution primitive of the spatial portfolio disclosed under USPTO provisional 64/049,409. Every other spatial primitive — marker-track transport, matched-pair settlement, n-party coordination, environmental disruption response — terminates in this gating step before any physical actuator commits.