

Evidence-Gated AI Authority

Slowing the expansion of AI authority without stopping the advancement of AI capability

Capability is not authority.

A system may evaluate, recommend, decide, or act. Each permission must be separately justified, scoped, current, and checked before consequential commitment.

The problem

AI-enabled workflows can produce fluent recommendations or trigger actions without making the path from evidence to judgment, uncertainty, authority, and outcome easy to inspect. Auditability after the fact is not enough if evidence coverage is unknown or permission is assumed from technical access.

The proposed operating pattern

1 · EVIDENCE	Record provenance, coverage, independence, conflict, and what remains missing.
2 · EVALUATE	Use explicit, versioned criteria and preserve result, method, and uncertainty.
3 · AUTHORIZE	Check authority for this actor and action at the point of commitment.
4 · ACT OR PAUSE	Execute only within granted scope; otherwise abstain, hold, or escalate.
5 · RECORD	Keep a reconstructable record of inputs, evaluation, authority check, decision, action, and escalation.

A concrete implication

If an invoice requests payment to newly emailed bank details that conflict with the vendor record, and independent confirmation is missing, the workflow should preserve that conflict, block the bank-detail change and payment release, and route the case to an authorized reviewer. Urgency does not supply missing evidence or create authority.

Where it could fit

Proposed control points are evidence/context adapters, a versioned evaluation orchestrator, a commit-time authority gate, bounded action adapters, and an audit/escalation sink. These could sit in an application service, workflow engine, or orchestration layer if all consequential paths pass through the controls and bypasses are tested.

Maturity and limits

ReasoningArc is a single-author, evolving research architecture. Current adversarial results are architecture-level tabletop evaluations, not runtime software tests, independent assurance, certification, regulatory approval, or proof of safety. The 12-problem taxonomy remains a hypothesis under active challenge. API/SDK contracts and a reference implementation are not specified in the current materials.

This companion executive brief summarizes the current working thesis. It is not the complete position paper, a product specification, or an implementation guarantee. Detailed architecture, test record, open questions, and change ledger: ReasoningArc.com.