



COMPANION NOTE · AGENT GOVERNANCE

The harness now enforces the policy. It still cannot tell you the policy is working.

Databricks moved agent policy out of the prompt and into the harness. That is the right layer. It is also where the governance theatre gap reappears.

Jason Cruz Falzon

Creator of the Cognitive Delivery Framework and the Cognitive Delivery Index
jason@cognitivedelivery.co.uk · cognitivedelivery.co.uk

A better control point, the same open question

Omnigent wraps every coding agent in one sandboxed session and lets you set policy at the layer above them. Prompt-only constraints were never governance. They were suggestions the model could talk itself out of. Moving the rule to where the action passes through is a real step.

The meta-harness announced at this year's summit gives teams one control point across Claude Code, Codex, and any framework they run. It carries three policy surfaces. Each one produces an artefact you can point to. None of them, on its own, tells you whether the policy changed what shipped.

That distinction is the whole argument. **Level 4** is a control you can document and show ran. **Level 5** is a control that demonstrably changes behaviour at the point of action. Omnigent gets you a cleaner Level 4 than the industry has had. The distance to Level 5 is still yours to close.

FIGURE 1 · OMNIGENT'S POLICY SURFACES, READ AGAINST THE LEVELS

POLICY SURFACE	WHAT IT PROVES (LEVEL 4)	WHAT LEVEL 5 STILL NEEDS
Cost budgets L4	A spend cap exists and the session paused when it hit the threshold.	Evidence the cap shaped real decisions, not a ceiling set so high it never binds.
Allow / deny / approve L4	A contextual rule fired and the action was blocked or routed for approval.	A deny rule that nothing triggers is documentation. Did it ever change an outcome.
Approval steps L4	A human approval event was logged against the action.	An approval cleared in seven seconds is a log line, not a judgement. Was it substantive.

GOVERNANCE THEATRE RISK

The harness can prove the policy fired. It cannot prove the policy was doing any work. A control that never binds, a deny rule nothing trips, an approval no one reads: all three log perfectly and all three are theatre. The cleaner the control plane, the more finished the work looks. It isn't.

Grade the governance, do not become it

The Cognitive Delivery Index does not enforce policy and does not want to. Omnigent and Unity AI Gateway enforce. The index reads what they produce and renders an independent judgement on whether the governance is substantive or theatre.

This is a layering point, not a competition. A meta-harness that defines enforceable policy at the harness layer is solving the enforcement problem well. The index solves a different problem: whether the enforced policy demonstrably changed behaviour, scored the same way across every platform an estate runs.

Why the judgement has to come from outside

The harness logging that its own rule ran is not evidence the rule mattered. A platform grading the governance it enforces is marking its own homework. For the AI Act, DORA, and ISO 42001, that is exactly the conflict an auditor discounts. An assessment is worth more precisely because the thing rendering it is not the thing doing the governing.

WHAT TO DO WITH THIS ON MONDAY

Take one team running agents through a harness. For each policy you have defined, find the last time it actually blocked, paused, or reshaped an action. If you cannot, the control is Level 4. Write down what Level 5 evidence would look like for that one policy, and instrument for it before you add another rule.

The six dimensions still apply

Provenance Integrity, Supervisory Clarity, Scope Definition, Review Substance, Constraint Fidelity, Capability Calibration. A harness improves the raw material for several of them. It scores none of them. The constraint that fires is Constraint Fidelity evidence only if it demonstrably held. The approval that logs is Review Substance only if the review was real.

The better the control plane gets, the more it looks like the work is done. Someone still has to grade whether the governance is real. Capability without governance is speed without control.