

The Cognitive Delivery Index.

A six-dimension instrument for assessing governed AI participation in software delivery. One sheet. One workshop. One honest read on where the team actually stands.

DIMENSION	L1 ABSENT	L2 INFORMAL	L3 DOCUMENTED	L4 COMPLETE	L5 SUBSTANTIVE	SCORE
01 Provenance Integrity Can you reconstruct what agents produced and why?	No record of agent activity exists. ☐	Logs exist but inconsistent across teams. ☐	Standard logging captured for agent runs. ☐	Every agent action is traceable end to end. ☐	Provenance actively informs review and decisions. ☐	SCORE
02 Supervisory Clarity Who is accountable for agent output, with what authority?	Accountability undefined when agents author. ☐	Implicit ownership, varies team to team. ☐	Named owner per agent workflow on paper. ☐	Authority and escalation paths are explicit. ☐	Owners actively exercise authority in practice. ☐	SCORE
03 Scope Definition Do agents operate within explicit boundaries?	Agents act with no defined limits. ☐	Boundaries live in conversation, not policy. ☐	Scope written but rarely consulted. ☐	Explicit guardrails enforced in tooling. ☐	Scope is tested, evolved, and defended. ☐	SCORE
04 Review Substance Is review of agent work substantive or procedural?	Output ships without meaningful review. ☐	Skim review, rubber-stamp common. ☐	Review process exists, applied unevenly. ☐	Reviewers trained for agent failure modes. ☐	Review catches what the agent cannot see. ☐	SCORE
05 Constraint Fidelity Are rules explicit enough for agents to follow?	Rules live only in human heads. ☐	Constraints in scattered, ageing docs. ☐	Rules captured in one place, not enforced. ☐	Constraints encoded for agent consumption. ☐	Agents reliably honour the constraints. ☐	SCORE
06 Capability Calibration Does expansion follow evidence or enthusiasm?	Adoption driven by hype alone. ☐	Anecdotal expansion of agent use. ☐	Use cases logged, success unmeasured. ☐	Evidence required before scaling further. ☐	Expansion paced by demonstrated reliability. ☐	SCORE

The shift from L4 Complete to L5 Substantive is where most organisations stall. Documented governance is not the same as governance that actually shapes decisions. This is the governance theatre gap — and closing it is the work.

↑ THE GOVERNANCE THEATRE GAP

MOST MATURE DIMENSION Which dimension is the team strongest on, and what makes that the case?	BIGGEST GAP Where is the distance between current state and target widest?	TOP PRIORITY – NEXT 90 DAYS One dimension. One commitment. Owned by one person.	SCORE PROFILE 01 · Provenance Integrity 02 · Supervisory Clarity 03 · Scope Definition 04 · Review Substance 05 · Constraint Fidelity 06 · Capability Calibration
---	--	---	---

Running the Index.

How to facilitate a CDI assessment that produces honest scores, real conversation, and a single committed action. Not a workshop. A diagnostic.

i BEFORE YOU BEGIN

WHO TO INVITE Five people from cross-functional perspectives — delivery, security, architecture, operations, and one engineer in the day-to-day. Not just engineers. Not just leaders.	WHAT TO SEND AHEAD The blank canvas, twenty-four hours in advance. Ask people to read the dimensions but not score until the room is together.	WHAT TO BRING Printed canvases. Pens. A timer visible to the whole room. Resist the urge to use a digital tool — paper makes the silence work.
--	--	--

ii THE SIXTY-MINUTE PROTOCOL

TIME	PHASE	WHAT HAPPENS	FACILITATOR NOTE
0–5	Frame	Set the purpose: an honest read on where the team stands. This is not a performance review. The goal is one priority for the next ninety days.	Name what good looks like at the end of the hour. One owner. One commitment.
5–15	Score in silence	Each person scores all six dimensions independently. No discussion. No looking across the table. Score the team as it is, not as it should be.	Silence is the assessment working. Protect it.
15–30	Compare	Reveal scores together. Map the distribution per dimension. Highlight the dimensions with the widest spread between high and low scorers.	The disagreements are the data. Spend time there, not on the consensus.
30–45	Diagnose	Discuss the shape of the profile, not the totals. What pattern emerges? Where is governance theatre likely hiding? Resist the urge to fix.	If two people score the same dimension differently, ask each what evidence anchors their score.
45–55	Commit	Pick one dimension. One person owns the next ninety days. One specific move that takes that dimension up by a level.	If you cannot pick one, you do not yet agree on the diagnosis. Go back to phase 4.
55–60	Close	Schedule the next assessment. Same room. Ninety days. Document the score, the priority, and the owner. End on time.	The next data point is the score in ninety days. Without that, you have had a conversation, not an assessment.

iii READING THE PROFILE · WATCH-OUTS

FOUR PROFILE SHAPES Flat & Low Spiky High & Flat Climbing	Systemic immaturity. Don't pick a dimension to fix. Pick a behaviour to change across all of them. Pockets of practice. Find what is working in the strongest dimension and spread the pattern, not the policy. Governance theatre risk. Stress-test every L4 claim. Ask what evidence makes this real, not documented. Balanced — and rare. The team has done genuine work. Pick the stretch dimension and push to L5.	WATCH-OUTS - The dominant voice scoring up. Make scoring silent. Reveal together. "We're definitely a four." Ask for the evidence. The evidence is the score. - Disagreement treated as failure. The variance between scorers is the most useful signal in the room. - Skipping the ninety-day commitment. Without it, you have run a conversation, not an assessment.
--	---	---

Three teams. Three shapes.

What different CDI profiles look like in practice — and the first move each team should make. Score totals tell less than score shapes.

CDI · V1.0

SPECIMEN PROFILES · 03 / 04

COGNITIVEDELIVERY.SUBSTACK.COM

SPECIMEN A

The Lagging Team

Early adoption, no scaffolding. Two engineers ahead, the rest behind.

CDI TOTAL

9 / 30

01 · Provenance Integrity		1
02 · Supervisory Clarity		1
03 · Scope Definition		2
04 · Review Substance		2
05 · Constraint Fidelity		1
06 · Capability Calibration		2

Two engineers have been using AI agents for three months. The rest of the team is curious but hasn't started. There is no agreed approach. Pull requests land with no indication of what was agent-generated. When something breaks in production, nobody can say why or how it got there.

FIRST MOVE

Establish provenance before scaling. Pause expansion of agent use until basic logging is in place. Provenance is the prerequisite for everything else.

SPECIMEN B

The Theatrical Team

Documented governance. Hollow practice. Looks ready on paper.

CDI TOTAL

22 / 30

01 · Provenance Integrity		4
02 · Supervisory Clarity		4
03 · Scope Definition		4
04 · Review Substance		3
05 · Constraint Fidelity		4
06 · Capability Calibration		3

Comprehensive AI usage policy. Documented review checklist. Quarterly audit. The team scores high on the canvas because the artefacts exist. None of them shape day-to-day decisions. Reviewers approve agent PRs they could not have written themselves. The audit never finds anything.

FIRST MOVE

Force Review Substance through to L5. Train reviewers specifically on agent failure modes. Until review is real, the rest is theatre.

SPECIMEN C

The Balanced Team

Quiet competence. Solid across the board, no standout strength.

CDI TOTAL

18 / 30

01 · Provenance Integrity		3
02 · Supervisory Clarity		3
03 · Scope Definition		3
04 · Review Substance		3
05 · Constraint Fidelity		3
06 · Capability Calibration		3

The team has done the work to reach Documented (L3) on every dimension. No glaring gaps, no standout strengths. The risk is not collapse, it is plateau. Without a forcing function, this team will stay at L3 across the board indefinitely.

FIRST MOVE

Pick the dimension closest to your business risk. Push it to L4. Use it as the template for the others. Don't try to lift all six at once.

READING NOTE

Specimen B scores higher than Specimen C on the total. Specimen C is the safer team to ship behind. The shape of a profile tells you more than its sum — and a high-flat shape is the most reliable indicator of governance theatre.

Levelling up, one move at a time.

What progressing from one CDI level to the next actually looks like in practice. One concrete move per transition. Not a roadmap — a starting point.

DIMENSION	L1 → L2 GET ON THE BOARD	L2 → L3 MAKE IT STANDARD	L3 → L4 MAKE IT COMPLETE	L4 → L5 CLOSE THE GAP
01 Provenance Integrity	Start logging agent use, even informally. A shared spreadsheet beats no record.	Standardise where logs live and what they capture. One source of truth, not three.	Make logs queryable and linked to commits, PRs, and incidents.	Reference logs in retros, reviews, and post-incident analysis. The provenance shapes decisions.
02 Supervisory Clarity	Name an owner per agent workflow. Even one name beats none.	Document an accountability matrix: who owns what, who escalates to whom.	Establish explicit escalation paths with named decision rights.	Owners actively use the authority. Decisions get made, not just documented.
03 Scope Definition	Articulate what agents shouldn't do. The negative list is faster to write than the positive one.	Capture scope per workflow in writing. Distinct from policy. Specific to context.	Encode constraints in tooling. Guardrails the agent literally cannot cross.	Test, evolve, and defend scope changes. Scope drift is the failure mode.
04 Review Substance	Require human review of all agent output. No exceptions, even for trivial changes.	Create checklists for agent PRs that surface the failure modes humans miss.	Train reviewers on agent failure modes. Subtle hallucinations. Plausible but wrong.	Reviewers add what the agent cannot see. Context, intent, second-order effects.
05 Constraint Fidelity	Capture rules in one shared place, even if scattered today. Inventory before reform.	Consolidate to a single source. Resolve conflicts. Retire dead rules.	Make rules machine-readable so agents can consume them as input, not as documentation.	Validate that agents honour rules in practice. Sample, audit, correct.
06 Capability Calibration	Track where agents are in use. You cannot pace what you cannot see.	Measure outcomes per use case. Quality and velocity, not just volume.	Set evidence thresholds for expansion. New use cases earn their place.	Pace adoption by demonstrated reliability, not by enthusiasm or competitive pressure.

The L4 → L5 column is where most organisations stall. The first three transitions can be project-managed. The fourth requires culture, repetition, and time. It is the most expensive work — and the most valuable.

↑ THE GOVERNANCE THEATRE GAP