

The Cognitive Delivery Index.

A six-dimension assessment of an organisation's readiness to deliver software in an age where humans and agents share the keyboard.

CDI · V0.1

ONE-PAGE REFERENCE

COGNITIVEDELIVERY.SUBSTACK.COM

i THE SPECTRUM

DEV 1.0 Solo Author Tools assist. Humans write. Agents are absent or invisible to the workflow.	DEV 2.0 Pair Coder Inline AI suggests. The human approves every line and remains the author of record.	DEV 3.0 Delegated Tasks Agents take defined slices of work. Humans set scope and review the output.	DEV 4.0 Orchestrated Agents Humans set intent and constraints. Agents plan, execute, and report against them.
---	--	---	---

ii THE SIX DIMENSIONS

01 Delivery Maturity Where the team operates on the Dev 1.0 to 4.0 spectrum. Whether agent use is incidental, integrated, or structural to the work. LOW AI use is personal, hidden, and inconsistent across the team. HIGH Agent involvement is explicit, documented, and reflected in process. DIAGNOSTIC Are we treating agents as power tools, or have we redesigned the work around them?	02 Governance Posture How agent activity is bounded, audited and controlled. The strength of the guardrails between intent and production. LOW No record of what an agent has changed, when, or why. HIGH Every agent action is traceable, attributable, and reversible. DIAGNOSTIC If an agent acted badly today, would we notice within minutes, weeks, or never?	03 Accountability Clarity Who owns the output when authorship is shared. Whether responsibility scales with autonomy or quietly evaporates. LOW "The AI wrote it" is treated as an explanation when something fails. HIGH A named human owns every shipped artefact, regardless of authorship. DIAGNOSTIC When something ships and breaks, can a named human still defend the decision?
04 Review Integrity Whether code review has been redesigned for non-human authorship, or whether reviewers apply human-author assumptions to agent output. LOW Reviewers skim agent PRs the way they always reviewed human ones. HIGH Review focuses on intent, edge cases, and failure modes, not surface correctness. DIAGNOSTIC Does our review process catch the failure modes agents actually produce?	05 Adoption Scale Whether hybrid delivery is isolated to pockets of practice or has become systemic across teams, functions and disciplines. LOW One or two enthusiasts on the team. The rest watch. HIGH New joiners learn hybrid delivery as the default way work is done. DIAGNOSTIC Could a new joiner work the way our best teams already do, or is the practice tribal?	06 Outcome Trust Whether the work produced can be maintained, explained and defended over time. Not just delivered quickly today. LOW Code ships fast. Nobody is confident maintaining it six months on. HIGH Agent-produced work is as durable, explainable and owned as anything else. DIAGNOSTIC If the agent disappeared tomorrow, would the team still own what it built?

iii USING THE INDEX

SCORING RUBRIC Each dimension is scored independently on a five-point scale. The composite gives a single read on readiness. The shape of the profile tells you where to invest. 0 ABSENT Not recognised as a concern. 1 AD-HOC Acknowledged but inconsistent. 2 DEFINED Documented practice exists. 3 MANAGED Practice is measured and improving. 4 SYSTEMIC Embedded in how the org operates.	RED FLAGS Cross-cutting warning signs that low scores are being masked by surface productivity gains. "We don't track that yet" said about agent activity. - AI usage growing faster than governance maturity. - Reviewers approving PRs they could not have written themselves. - Productivity claims without paired trust metrics. - Pockets of brilliance with no organisational pattern. - Velocity rising while incident rates rise alongside.	RUN AN ASSESSMENT A 60-minute structured self-assessment with the right people in the room. - Convene a cross-functional group. Delivery, security, architecture, ops. Not just engineers. - Score each dimension independently. No discussion until everyone has scored alone. - Compare and discuss the gaps. The disagreements are more useful than the averages. - Act on the lowest two. The highest does not need investment. Trends matter more than scores.
---	---	---