

COGNITIVE DELIVERY / PAPER

Proximity is calibration

Why staying close to the work stopped being a preference and became a governance control.

Jason Cruz Falzon

Creator of the Cognitive Delivery Framework and the Cognitive Delivery Index

Why this paper exists

A conversation is running through delivery organisations about senior people going back to hands-on work. It is usually framed as appetite. That framing is comfortable, and it is wrong in a way that matters.

The appetite was always there. Senior people have always missed building. What changed is not desire but permission. Staying close to the work used to read as a failure to delegate. It now reads as founder mode. Same behaviour, opposite status signal.

Underneath the change in fashion sits a real problem, and the coverage keeps walking past it.

As AI moved the work from producing to directing, it also removed the thing that produced judgement in the first place. Proximity to the artefact was never only about output. It was how people stayed qualified to evaluate output. That was a by-product, it was free, and because it was free nobody counted it as a cost when they took it away.

This paper argues that staying close to the work is a control. Not nostalgia, not a productivity choice, and not a personality trait. A control, in the sense a regulator would recognise.

RULE 01

Your approval is worth the distance between you and the thing you approved.

How to use this paper

Sections one to three set out the mechanism. Section four connects it to an obligation deployers already carry. Section five is a diagnostic you can run this week without instrumentation. Section six states the objection I find hardest to answer, because a paper that only argues one side is an advert.

Everyone moved up a layer

The division of labour in software delivery changed over the last eighteen months, and it changed without anyone taking a decision.

AI now writes a substantial share of production code. Sonar's 2026 survey of more than 1,100 developers puts it at 42% of committed code, with respondents expecting 65% by 2027. Sonar sells code review and verification tooling, so the finding is not disinterested. Treat the figure as directional. The direction is not in dispute.

What follows from that volume is a quiet reshuffle of who does what.

FIGURE 1 / THE SHIFT NOBODY PROPOSED

LAYER	WAS DOING	IS NOW DOING
Engineer	Writing the change	Reviewing a generated change
Lead	Reviewing the change	Reviewing the review
Manager	Reviewing the review	Approving a summary
Director	Approving a summary	Reading a dashboard

Each of those moves is defensible on its own. Every one of them looks like sensible delegation, and in isolation every one of them is.

Stack four of them and you have built something else. The distance between every approval and the artefact it approves has grown by one layer, everywhere, at the same time.

Nobody proposed this. It emerged, which is why there was no point at which anyone was asked to accept the trade.

That is the part worth sitting with. If no one decided to move each approval a layer further from the artefact, then no one assessed what the added distance costs. The change arrived as a series of local efficiencies, and local efficiencies do not get risk assessed. They get celebrated in a retrospective and absorbed into the baseline.

The rest of this paper is about what that distance does, and why it is a governance question rather than an operating model one.

Verification is not review

The load-bearing claim in this paper is about cognition, not org design. It is that checking work and doing work are different acts, and only one of them builds understanding.

DORA states it plainly: verification is a fundamentally different cognitive task than creation.

When you write something, understanding arrives with the act of writing. You cannot produce the thing without holding a model of it in your head. The comprehension is free, because it is a by-product of the work.

When you inherit something, no comprehension arrives. You have to construct it deliberately, from the outside, against the clock, for a change you did not choose. Werner Vogels calls the resulting cost verification debt, which is a good name for it.

That asymmetry is not new. Reviewing another person's code was always harder than writing your own. What changed is the volume and the provenance, and one engineer in DORA's study put the consequence better than any analyst has:

Reviewing another's code is so much harder than writing it. AI tools are increasing the rate at which people can churn out code that needs to be reviewed.

Engineer quoted in DORA, March 2026

Read that as an accounting statement. The gain accrues to the author. The cost lands on the reviewer. That is a transfer, not a saving, and it is invisible in every metric that counts throughput.

The Sonar data points the same way from a different angle. 38% of developers say reviewing AI-generated code takes more effort than reviewing a human colleague's, and while 96% report not fully trusting AI output, only 48% always check it before committing. Vendor caveat applies again. The gap between what people believe and what they do is the part worth keeping.

ON THE EVIDENCE

The DORA analysis draws on 1,110 open-ended responses from engineers at a single company, published March 2026. It is qualitative, it is five months old, and Google Cloud runs DORA. I use it because the mechanism it describes is structural rather than local, and because the finding runs against the interest of the organisation publishing it.

What proximity was quietly producing

Here is the part the flattening conversation misses. Proximity to the work was never only about producing work. It was the mechanism by which people became able to judge work.

You saw enough changes up close to develop a sense of what good looked like. You watched enough things fail to recognise the shape of a failure before it fully arrived. Nobody designed that and no budget line carried it. It was a by-product of being near the artefact.

Remove the proximity and the by-product goes with it.

DORA names this the expertise paradox: AI lets an engineer work productively in domains where they lack the context to judge correctness. One participant described the result as the blind leading the blind. The report goes further, finding that these usage patterns deliver breakthrough productivity while simultaneously blocking skill development, limiting the hands-on problem solving that apprenticeship depends on.

Now scale that upward.

If the engineer cannot fully judge the AI output, and the lead cannot fully judge the engineer's review, and the manager cannot fully judge the lead's summary, then the organisation holds a chain of approvals in which no single link is independently calibrated.

Every signature in that chain is real. Every one of them was given in good faith. The aggregate assurance is still theatre, and an audit will find nothing wrong with it, because the thing that is missing was never written down as a requirement.

WHAT GETS MEASURED

- The review happened
- The approver was named
- The gate was passed
- The policy exists and is current
- Throughput rose

WHAT ACTUALLY GOVERNS

- The reviewer could have caught it
- The approver had seen comparable work recently
- The gate has refused something
- The policy changed a decision this quarter
- Somebody can still say the summary is wrong

The left column is documentation. The right column is governance. Most organisations can produce the left on demand and have never been asked for the right.

The regulator has already said this

The most precise statement of this problem I have read this year came from a compliance practitioner rather than an engineer, and it fits in a single line.

An organisational chart is not an oversight control.

IAPP, deployer readiness analysis, May 2026

The assessor writing it had run readiness work across around a dozen organisations, and the pattern was consistent: asked to evidence human oversight, teams point at a reporting line and assume that settles it.

Article 26(2) of the EU AI Act requires deployers of high-risk AI systems to assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support.

Competence is the word carrying the weight, and it is doing something specific. The Act does not ask who is nominally accountable. It asks whether the named person was in a position to intervene. The evidence that satisfies it is the named reviewer, the decision authority, the intervention trigger, the escalation route, and the record of what happened when the system behaved abnormally.

That is the same argument this paper has been making, arrived at independently, from the opposite direction, in statutory language.

A note on timing, because much of the commentary is now stale

The AI Act became applicable on 2 August 2026 and the Commission began enforcement the same day. The AI Omnibus entered into force on 27 July 2026, moving the high-risk obligations for Annex III systems to 2 December 2027 and for systems embedded in products to 2 August 2028.

What moved is when enforcement starts. Article 26 itself is unchanged. Analysis published earlier in 2026 against an August 2026 high-risk deadline should be read with that in mind.

The practical reading is the same one I would give regardless of the date. Anchor to the obligation, not the deadline. The obligation is that oversight must be real, and an extra sixteen months does not make a reporting line into a control.

The proximity test

Five questions. Answerable this week, without instrumentation, tooling or a maturity assessment. They are diagnostic rather than scored, and the answers tend to be uncomfortable in a useful way.

01 When did you last read an artefact rather than a report about one?

Not a dashboard and not a summary. The actual change, specification or decision record. If the honest answer is measured in months, your sense of what good looks like is historical, and it is ageing faster than it used to.

02 Could your named approver have caught it?

Take one recent approval on something that mattered. Ask what the approver would have had to see in order to catch a serious defect, and whether they saw it. This is the Article 26 question in operational clothing, and it is the one most organisations cannot answer.

03 What happens to a review that finds nothing?

If a thorough review and a cursory one leave identical traces in your evidence, then what you are recording is that reviews occurred. You are not recording that they bit. Those are different claims and only one of them is a control.

04 Where does the verification cost appear in your plan?

If throughput rose and no line absorbed the added review load, the cost did not disappear. It was absorbed silently by whoever reviews, which is usually your most experienced people, which is usually the constraint you were trying to relieve.

05 Who is being trained to judge?

If AI now does the work that used to build judgement, name the mechanism that replaces it. If you cannot name one, you do not have a tooling problem. You have a bench problem arriving in about three years, and the people who would have filled it are being hired away from it now.

None of these require you to be hands on in the way an engineer is hands on. They require you to be close enough to the artefact that your judgement about it is current.

The objection I find hardest

A paper that only argues one side is an advert, so here is the strongest case against my own position.

Being hands on at a senior level has a real failure mode. You become the bottleneck on the interesting problems. You take the work that would have grown the people below you, which is precisely the harm this paper claims to be worried about. And you can describe that as calibration for years without anyone challenging you, because the story is flattering and the alternative is admitting you would rather be building.

Staying close to stay calibrated, and staying close because it is more enjoyable than the management work, feel identical from the inside. I have not found a clean test. The only honest one I have is directional: are you taking work off your team, or sitting alongside it?

The second objection is empirical. The flattening story is thinner than the coverage suggests: Glassdoor has younger workers reaching management on schedule despite the unbossing rhetoric, and moving into management still paid better than staying put. Stated preference is not behaviour. This paper is about a mechanism rather than a trend, because the mechanism is better evidenced and will outlast it.

Closing

There is a version of the next few years in which organisations become flatter, faster and more productive, and quietly lose the ability to tell whether any of it is correct.

The signatures will all be in place. The controls will all be documented. Somewhere in the middle of it the last person who could have said the summary was wrong will have moved up a layer, and there will be no incident, no alert and no finding to mark the moment it happened.

How far are you, right now, from the last thing you approved?

SOURCES

SOURCE	WEIGHT
DORA (Google Cloud), Balancing AI tensions, 10 March 2026	Primary. Qualitative, single company, five months old. Finding runs against publisher interest.
Sonar, 2026 State of Code Developer Survey, 8 January 2026	Vendor-run. Sonar sells verification tooling. Directional only.
IAPP, EU AI Act deployer evidence gaps, 8 May 2026	Practitioner analysis, around a dozen readiness assessments. Predates the Omnibus timeline change.
European Commission, AI Act framework page, updated 3 August 2026	Regulator. Used for every date claim in this paper.
Business Insider, on Philip Su, 19 February 2026	One practitioner essay reported as news. Argument, not evidence.
Forbes reporting Korn Ferry Workforce 2025	Contributor piece, figures quoted second hand. Trace before reuse.

Written in a personal capacity. No client or engagement material is used anywhere in this paper. Correspondence to jason@cognitivedelivery.co.uk