// WHITE PAPER · PUBLISHED 16 SEPTEMBER 2026

AI and Headcount: Somebody on the Board Did the Arithmetic

What the evidence supports, what it complicates, and the sequence that keeps the decision reversible

Part 9 of 10 in *The Money Language Series*. Each piece stands on its own. The others: Part 1 · Part 2 · Part 3 · Part 4 · Part 5 · Part 6 · Part 7 · Part 8 · Part 10.

Jason Newell · MAX Research Collective · Report WPR-BIZ-010 · Published 16 September 2026
Licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/
Original: https://www.maxresearchcollective.com/articles/ai-and-headcount.html
Cite as: Newell, J. (2026). AI and Headcount: Somebody on the Board Did the Arithmetic. MAX Research Collective, WPR-BIZ-010. https://www.maxresearchcollective.com/articles/ai-and-headcount.html

Somebody on your board has done the arithmetic. If AI makes engineers 30 percent more productive, and you have 60 engineers, that is 18 engineers of capacity you are paying for and do not need.

The arithmetic is clean. Every input is contested. How much capacity did you actually gain, and where in the pipeline did it land?

This topic attracts two useless positions: AI changes everything and headcount is a legacy concept, or AI is autocomplete with a marketing budget. The evidence supports neither position, and a technology leader still has to decide something real in the middle of them.

What the measurement shows

DORA's 2025 report drew on nearly 5,000 technology professionals and more than 100 hours of qualitative work. It found roughly 90 percent AI adoption, up 14 points year over year, with median usage around two hours a day (about a quarter of a

working day, every day); over 80 percent of respondents believed AI had increased their productivity. Belief is not measurement. Hold that thought.

The same research found something more useful than a productivity percentage: a reversal. In 2024, AI adoption correlated with reduced delivery throughput, which was the year the industry spent explaining that the tools were new and the measurement was early, and in 2025 the sign flipped, so adoption now correlates positively with throughput and with product performance. Across both years, though, adoption correlated with increased delivery instability, which shows up as change failures and as longer time to restore.

DORA's framing is that AI is an amplifier. It magnifies what is already there: the strengths of a high-performing organization, and the dysfunctions of a struggling one. Neither gets a vote. Faster code generation exposes whatever sits downstream of it. Where review and testing were already the constraint, added capacity upstream floods that constraint rather than relieving it.

The trial that gets weaponized

METR ran a randomized controlled trial with 16 experienced open-source developers across 246 real tasks in their own mature repositories, projects they had worked on for an average of five years. Each task was randomly assigned to allow or disallow early-2025 AI tools, mostly Cursor Pro running Claude 3.5 and 3.7 Sonnet.

Before starting, the developers forecast that AI would make them 24 percent faster. After finishing — after doing the work, on their own code, with the tools in front of them — they estimated it had made them 20 percent faster. Measured, they were 19 percent slower. That is a 39-point swing between the felt number and the recorded one. The perception gap, rather than the slowdown, is the finding worth carrying out of the study.

Now the caveats, in full, because this result gets used as a weapon. Sixteen developers is small. The tooling predates the agentic tools that dominate now. The setting was deep familiarity with a mature and complex codebase, which is close to the hardest case for AI assistance: the participants already held in their heads most of what a model would have to be told, so the tool spent its time being brought up

to speed on a system its user had known for five years. METR itself labels the result historical and says it does not reflect current tools. Its follow-up cohort with agentic tooling came in around 4 percent slower — the interval runs from roughly 15 percent slower to 9 percent faster — so a real speedup sits inside the range. METR's own read is that developers are more sped up in early 2026 than they were in early 2025, with selection effects leaving the magnitude uncertain. And about 69 percent of the original participants kept using AI tools after the study ended. Why would they, if the tools cost them time? That question has no clean answer in the data, and it is the honest end of the caveat list.

Set against all that, a controlled trial of GitHub Copilot on a standardized, self-contained task found participants finished about 56 percent faster, with the largest gains among less experienced developers.

SETTING	WHAT THE EVIDENCE SUGGESTS
Well-scoped, self-contained tasks	A large and real speedup
An unfamiliar codebase or language	Meaningful help
Boilerplate, scaffolding, tests, glue	Strong
Deep changes in a complex system you know well	Marginal, possibly negative
Less experienced developers	Larger gains
Highly experienced developers on their own code	Smallest gains

The gains are real, they are larger than they were a year ago, and they are distributed unevenly across the work your team actually does. A headcount plan built on one organization-wide multiplier is built on an average that describes nobody.

What happened to the companies that cut first

Here the evidence is clearer, and less flattering.

Klarna said publicly that its AI assistant did the work of roughly 700 agents and handled about two-thirds of customer service chats, which is the single most quoted number in this whole debate and the one most often quoted without its sequel. It froze non-engineering hiring and let headcount fall substantially. Then the CEO acknowledged that the cost-cutting had gone too far and that service quality had suffered. The company is hiring human agents again, and describes investment in the quality of human support as the direction going forward.

Orgvue surveyed 1,000 C-suite and senior decision-makers at medium and large organizations. Thirty-nine percent had made redundancies after deploying AI. Of those, 55 percent said they had picked the wrong roles. In the same research, 34 percent reported employees resigning as a direct result of AI deployment; 80 percent planned reskilling programs.

Gartner projects that half of the organizations that reduced customer service headcount because of AI will rehire for similar roles by 2027. That is one in two, forecast by the firm those organizations pay for forecasts.

Careerminds surveyed 600 HR professionals who had made layoffs in the previous year. About a third had already rehired a quarter to a half of the eliminated roles, and another third had rehired more than half. Over half of them rehired inside six months. Over half said AI required more human insight than they had anticipated; only about 21 percent said AI fully replaced roles with no operational issues.

Four independent sources point the same way. The shared finding is narrower than "AI does not work": organizations systematically overestimated which specific roles could be removed, discovered the judgment those roles were carrying, and paid to rebuild that capacity inside a year.

Rehiring within six months costs more than never cutting; the arithmetic is not close. You pay severance, lose institutional knowledge and damage the employer brand, then pay recruiting fees and ramp time for somebody who knows less than the person who left.

The bottleneck moves

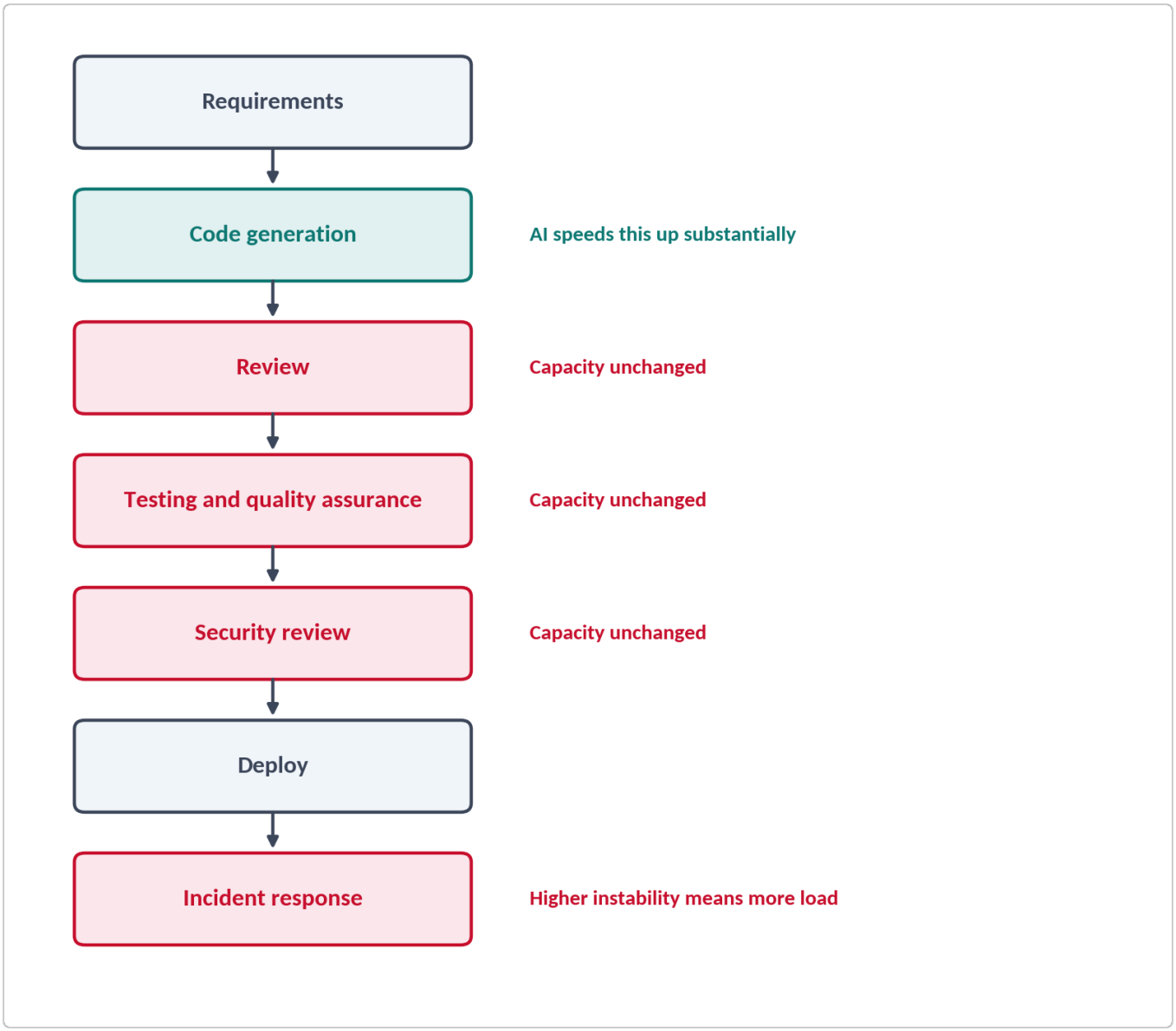


Figure 1. One stage got faster and six did not. The green stage is the one AI accelerates, the red stages are the ones that absorb the result, and the last of them is where an unabsorbed queue eventually shows up.

This is the operational consequence of DORA's instability finding, and it is the most practical thing in this article. Generate code faster while every downstream stage keeps the same capacity and you have built a queue. Where does a queue like that drain? It drains as incidents.

So the correct first investment, when an organization adopts AI seriously, is usually downstream capacity: test coverage, faster CI and review time. DORA's own

conclusion says the same thing from the other direction: AI value comes from the quality of internal platforms and workflows rather than from the tools.

One more finding from the same period belongs here. Developer trust in AI output is low, and Stack Overflow's 2025 survey found around 46 percent actively distrusting AI accuracy against 33 percent trusting it, with about 3 percent reporting high trust. People are already doing the verification (which is the sane response to a tool you distrust). Time saved in generation is being reallocated to auditing, which is exactly what DORA's qualitative work describes.

The pipeline problem nobody budgets for

Juniors learn by doing the work AI now does well: boilerplate, simple bug fixes, small features. That is the apprenticeship, and it is the cheapest capacity a team ever builds.

Stop hiring juniors because AI covers that tier and the cost is deferred rather than removed. In five years the senior bench is whoever you managed to poach, at market rates set by everybody else who made the same decision.

Early research suggests AI assistance can impair conceptual understanding and debugging skill when it substitutes for the productive struggle that builds expertise. It is early work (small samples, short horizons) and it deserves a light touch. It points the same way as the structural argument, and the structural argument stands on its own.

So: hire juniors, and change what they do. The apprenticeship shifts from writing simple code to reviewing generated code and debugging systems they did not write. That is arguably a better education for the work that is arriving, and it is a reason to redesign the first two years rather than to skip them.

A defensible way to decide

Industry productivity multipliers belong nowhere near a headcount plan, because there is no single number to apply. Run this sequence instead.

STEP	WHAT YOU DO	HOW LONG
1	Baseline throughput, change failure rate, lead time, review latency and incident volume	Already running, ideally
2	Adopt the tooling properly: training and patterns, rather than licences alone	One quarter
3	Invest in the downstream constraint first: tests, CI, review capacity	Concurrent
4	Hold headcount flat and measure what the same team ships	Two quarters
5	Read the whole scorecard: throughput, stability, rework and burnout together	End of period
6	Then decide: more roadmap, reallocation, slower hiring, or reduction	

Slowing hiring differs from cutting in the one respect that matters most: it is reversible. Given that roughly 55 percent of the leaders who cut said they got the roles wrong, reversibility carries real option value.

If you do reduce, reduce against demonstrated capacity in your own codebase rather than against a benchmark, and know which specific knowledge walks out with each person. The Careerminds data says a third of organizations named loss of critical skills as the consequence they actually experienced.

What I actually do

1. Treat AI as a capacity multiplier before treating it as a headcount reducer. The sequence is to ship more roadmap, measure what changed, and only then decide.
2. Fund the downstream capacity before the tooling or alongside it, which means review, test and observability.
3. Measure against the full DORA scorecard rather than one row of it. Throughput on its own will tell a flattering lie for about two quarters, and then stop.

4. Distrust self-reported productivity, including my own. The perception gap is the best-replicated finding in this literature, and it applies to whoever is reading the dashboard as much as to whoever is writing the code.
5. Keep hiring juniors, and redesign their first two years around reviewing and debugging what the tools produce.
6. Prefer slowing hiring to cutting, because it is reversible and far cheaper when it turns out to be wrong.
7. Keep every number in a headcount plan traceable to something measured inside the company.
8. Say the quiet part to the board. AI is delivering real gains, smaller and later than the headlines, and the companies that cut hardest are currently rehiring. That is defensible, it is evidence-backed, and it is more useful to a board than optimism.

Glossary

TERM	DEFINITION
Agentic tools	AI systems that take multi-step actions with more autonomy than autocomplete
Amplifier thesis	DORA's finding that AI magnifies existing organizational strengths and dysfunctions
Change failure rate	The share of deployments that cause a degradation
Confidence interval	The range a measured effect plausibly falls within
Delivery instability	Change failure rate together with time to restore service
DORA	DevOps Research and Assessment, the long-running delivery performance research program
METR	Model Evaluation and Threat Research, a nonprofit AI evaluation organization
Perception gap	The difference between how fast people feel and how fast they are
Randomized controlled trial	The strongest common study design for a causal claim
Selection effect	Bias arising from who chooses to take part in a study
Throughput	How much change you ship, and how fast

References

- DORA, 2025 State of AI-assisted Software Development, Google Cloud, September 2025

- DORA, "Balancing AI tensions: Moving from AI adoption to effective SDLC use"
- IT Revolution, "AI's Mirror Effect: How the 2025 DORA Report Reveals Your Organization's True Capabilities"
- Becker, Rush, Barnes and Rein, Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity, METR, July 2025, arXiv 2507.09089
- METR, "We are Changing our Developer Productivity Experiment Design", February 2026
- Orgvue, annual workforce research, April 2025
- Gartner, customer service rehiring projection, 2025
- Careerminds, survey of 600 HR professionals, February 2026
- Forrester, Predictions 2026
- Stack Overflow, 2025 Developer Survey, for the AI trust data
- Bloomberg interviews with Klarna CEO Sebastian Siemiatkowski, reported via Tech.co, eMarketer and Maginative

Best next read

Part 10, Outsourcing After Section 174A. Headcount decisions and location decisions are entangled now in a way they were not two years ago, because the tax code started treating domestic and foreign development very differently. It closes the series.

The full series

PART	TITLE
1	CapEx, OpEx, and EBITDA: The Same Dollar, Two Different Stories
2	Reading the Financials: The Primer for Everyone Who Never Had To Care
3	Capitalize or Expense: Your Jira Board Is Now Audit Evidence
4	Depreciation and Useful Life: The Chip Outlives Its Economics
5	Placement Economics: The Right Answer, Applied Everywhere
6	Lift-and-Shift or Re-Architect: The Two That Modernized Later
7	Vendor Economics: The Email in November
8	AI Spend: Three Budgets Pretending to Be One
9	AI and Headcount: Somebody on the Board Did the Arithmetic, which is this one
10	Outsourcing After Section 174A

This series is an introduction for technology leaders. It is accounting, tax and workforce context rather than accounting, tax or employment-law advice. Treatment is fact-specific. Model it with your CFO, your tax advisor and your HR leadership before you commit.

CAREER & LEADERSHIP WPR-BIZ-010

// ABOUT THE AUTHOR

Jason Newell leads organizations and people in technology, and is a researcher, deep technologist and builder — from a CRAY to agentic AI. He founded MAX Research Collective to build AI for regulated industries that runs where its owner has control, with healthcare as its strongest focus. About Jason → Connect on LinkedIn →

Newell, J. (2026). MAX Research Collective, WPR-BIZ-010. CC BY 4.0