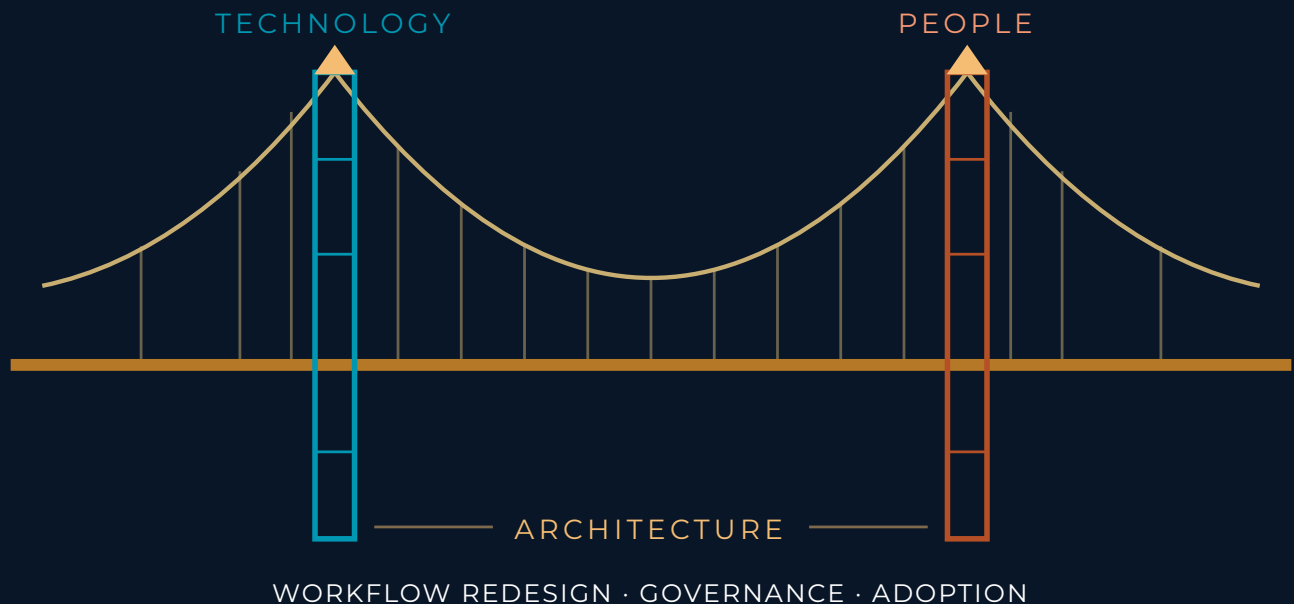


What Actually Returns

A model for AI and people investment, built on published evidence.



The two towers of AI value, the architecture that spans them, and a worked example for a 100-person company.

Naomi Withers
July 2026 · v1

A defensible range, and what it shows

Most AI business cases rest on a single confident number that no one can source. It looks precise, it goes into a board slide, and it collapses the moment someone asks where it came from. This paper offers the alternative: a defensible range built on published evidence, and a model that shows what AI and people investment actually return when they are built together.

The argument in six points:

- 01 The technology works, but the return has a ceiling.** Controlled experiments show double-digit productivity gains from AI. Yet only about 39% of organizations report any enterprise-level earnings impact, and roughly 3% report a return in the 10–20% range. The gap between the lab and the profit-and-loss statement is the real story.
- 02 Most of the return leaks out before it reaches the income statement.** Nearly two-thirds of organizations have not begun scaling AI, and the large majority of pilots stall before measurable financial impact. The stall happens at integration, where a capable tool never becomes the workflow the company runs on.
- 03 The return on AI is really a return on your people.** AI compresses the skill distribution, lifting the least experienced most. The same tool produces different results in different organizations, and the difference is the workforce and the managers guiding it.
- 04 Human capital is the largest lever most mid-market companies own and the one they measure least.** Engagement, coaching, and leadership development have peer-reviewed returns large enough to dwarf most software budgets, once you put a number on them.
- 05 The two towers carry more together.** For a 100-person company, the technology tower and the people tower together model a structurally supported value of about **\$1.06 million a year** — more than either tower alone, because each strengthens the other.
- 06 The difference between what AI could return and what most companies get is an architecture gap.** It is made of workflow redesign, governance, and adoption. It is measurable, and the mid-market is unusually well positioned to close it.

Every figure in this paper traces to a graded, published source, listed in the evidence appendix. Where the causal proof does not yet exist, the model says so and holds back. That restraint is the point.

The problem with the number in your board deck

Most leaders I work with have already been handed a number. A vendor deck promised a 40% productivity gain, or a respected peer quoted a figure at a conference that sounded close enough to true. The number goes into a board slide. Then it sits there, waiting for someone to ask where it came from.

When that question comes, most AI business cases fall apart. Not because AI does nothing, but because the number was never sourced, and a number you cannot defend is worse than no number at all.

The honest version is a range, and it comes with its sources attached. Building that honest version reveals something the flattering one hides: most of the return a company is missing is not locked inside the technology. It is locked behind an architecture the company has not built yet. To see why, take the two towers in turn.

THE FIRST TOWER

What the technology actually returns

When researchers run real experiments, with random assignment and measured output, the productivity gains from AI are genuine and, in places, large.

Across three field experiments at Microsoft, Accenture, and a Fortune 100 firm, developers using an AI coding assistant completed **26% more tasks**. Among customer-support agents, AI assistance raised issues resolved per hour by **15% on average, and 30% for the least experienced**. Professional writers finished tasks about **40% faster**, and the weakest writers improved the most. Management consultants working inside AI's current capability produced work rated substantially higher, while doing measurably worse on tasks that sat outside it.

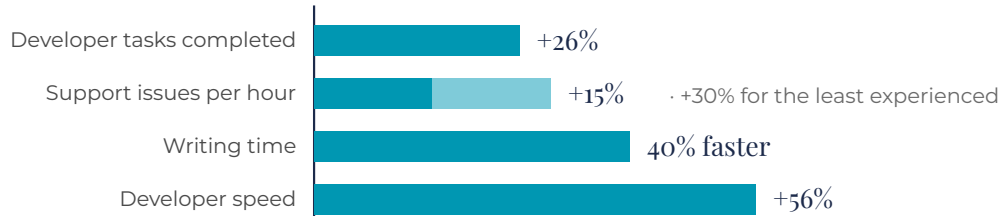
Two things run through all of it. The gains are real. And they concentrate at the bottom: AI narrows the distance between an organization's strongest and weakest performers. Hold onto that second finding. It quietly turns AI from a technology story into a people story.

FIGURE 1

Large at the task. Thin at the profit-and-loss statement.

IN CONTROLLED STUDIES

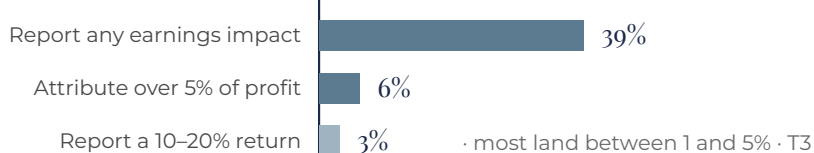
Measured gain on the task, per experiment



T1 · randomized and field experiments, 2023–2025

AT THE ENTERPRISE LEVEL

Share of organizations, per survey

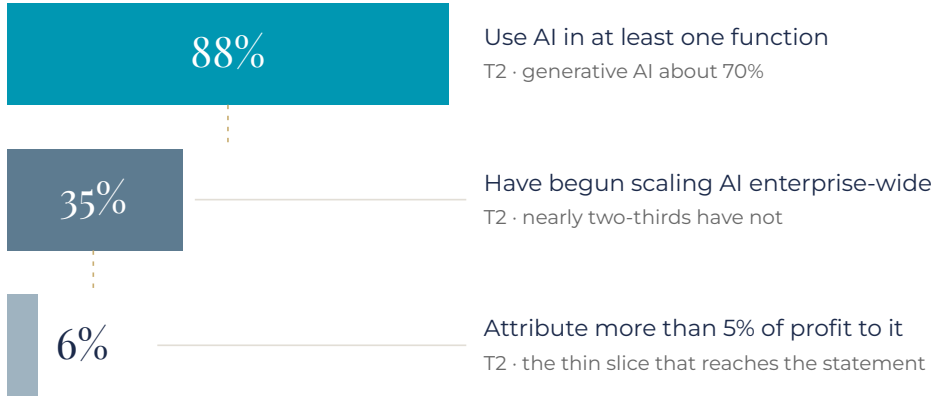


The two panels measure different things — a gain per task above, a share of firms below — and are shown together because the distance between them is the finding. Sources: Appendix B, Workforce and Competitive.

Now the number the vendor decks leave out. Of the organizations investing in AI, only **39% report any impact on earnings at the enterprise level**, and roughly **6%** attribute more than five percent of profit to it. A separate survey found only **3% of companies report a return in the 10–20% range**; most land between 1 and 5%. Controlled studies show double-digit gains on individual tasks. Company-level results show a thin slice of firms capturing real profit. The distance between the lab and the profit-and-loss statement is the whole story of AI right now.

That distance has a shape. **Nearly two-thirds of organizations have not begun scaling AI** across the enterprise. Pilots run, demos impress, and the work stalls before it reaches the income statement. One widely cited and openly debated analysis put the share of enterprise generative-AI pilots with no measurable profit impact at **roughly 95%** — a figure to treat as directional rather than precise, because the methodology is contested, though the direction is not. The failure mode is not model quality. It is the integration gap: the distance between a tool that can do a task and an organization redesigned so that it does.

FIGURE 2
The integration gap, in three steps



The failure is not model quality. It is the distance between a tool that can do a task and an organization redesigned so that it does.

Three separate surveys with different bases, not one cohort tracked through three stages. Shown to indicate shape. Sources: Appendix B, Tech / Data and Competitive.

There is a hopeful number in the same research. Top-performing mid-market firms scaled a working pilot in about **90 days**, while large enterprises took nine months or more. Smaller companies are often more able to close the lab-to-P&L gap, because fewer layers sit between the decision and the desk.

The most undermeasured number in business

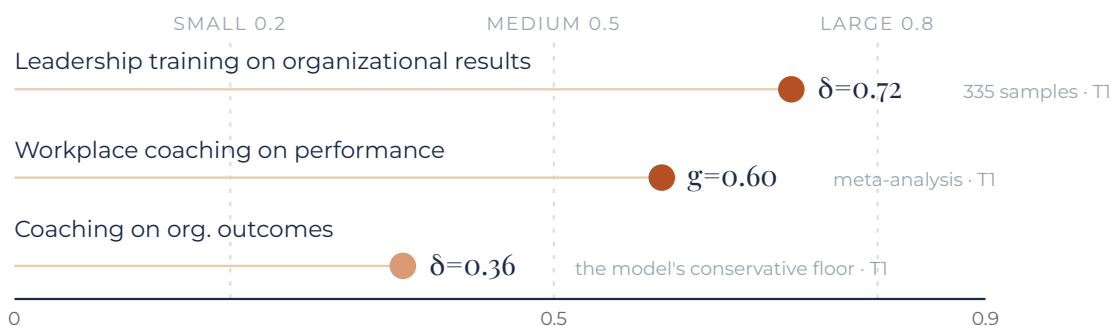
When a quarter gets tight, the people investment is the first line cut. The coaching budget, the development program, the engagement work. They go first because no one has put a dollar on them, and a number nobody can defend is easy to erase. That is the quiet problem with human capital. It is the largest lever most mid-market companies own, and the one they measure least.

Start with engagement, because it sets the ceiling on everything else. Only **20% of employees worldwide are engaged** at work, the second straight annual decline. Low engagement costs the global economy around **\$10 trillion a year**, roughly 9% of GDP. For a 100-person company, a 10-point engagement gain models to about **\$180,000 a year** in recovered performance, using the 18% productivity premium engaged teams show. The money is not missing. It is sitting inside disengagement.

Human capital ROI (return on investment) stops being a slogan when you look at the effect sizes. Workplace coaching has a peer-reviewed effect on performance of about **$g=0.60$** across meta-analyses. Leadership training moves organizational results at **$\delta=0.72$** across 335 samples. And **70% of the variance in a team's engagement** traces to one person: the manager. These are effects established across hundreds of studies in peer-reviewed journals.

FIGURE 3

The undermeasured lever, on a standardized scale



Standardized effect sizes, not percentages: how far the average outcome moves, in standard deviations. For scale, an effect of 0.60 puts the average coached manager above roughly three-quarters of the uncoached.

Sources: Lacerenza et al. 2017 · Theeboom et al. 2014 · Jones, Woods & Guillaume 2016. Appendix B, Leadership and Workforce.

FIGURE 4

The lever sits with one person

70%

of the variance in a team's engagement traces to its manager.

T2 · GALLUP, UPDATED 2026

75%

of HR leaders say their managers are overwhelmed.

T2 · GARTNER, 2024

The largest people lever and the most overloaded role are the same role. That is the whole case for funding managers before programs.

Turned into dollars for the 100-person company: coaching ten managers returns about **\$173,000 a year**, a figure the model has already halved for attribution honesty, netting a \$60,000 investment to roughly **\$113,000 a year**. Retention is larger still — five points of lower attrition is worth about **\$375,000** at a replacement cost of three-quarters of salary. That single line dwarfs most mid-market software budgets.

And notice what the model does with the figures it is least sure of. It credits no return at all to generic wellness programs, because two large randomized controlled trials found no significant effect on spending, absence, or performance. It holds financial-wellbeing benefits to a small, capped figure it labels directional, because the clean causal proof does not exist yet. That is not a verdict that financial wellbeing fails to pay; it is a measurement problem. Financial habits change slowly, and the longer studies that would capture the return have not been run. The model will not credit a long-game return it cannot yet source, and that restraint is exactly why the numbers above it can be trusted.

WHAT THE MODEL REFUSES TO COUNT

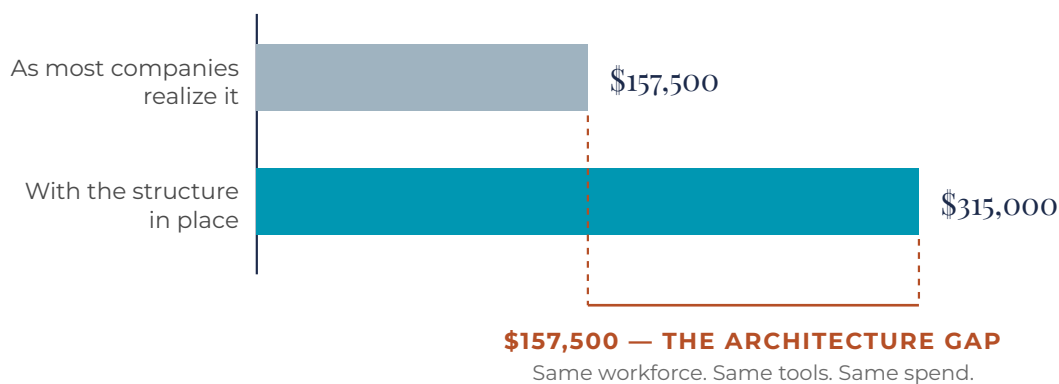
Generic workplace wellness earns a zero, on the evidence of two large randomized controlled trials. Financial wellbeing is carried at a small gross figure and flagged directional. Both decisions cost the model money on paper. Both are why the rest of it holds.

The architecture gap

Across both towers, a pair of numbers keeps surfacing. For the 100-person company, the technology tower returns about **\$157,500 a year as most companies realize it**, and about **\$315,000 with the structure in place**. Same workforce, same tools, same investment. The difference — **\$157,500 a year** — is not a technology gap. It is an architecture gap.

FIGURE 6

The architecture gap on the technology tower



100-person company, base uplift scenario. Source: the Living ROI Calculator, changeadvisor.net.

The model applies a 50% discount to the AI value by default, and it is not being pessimistic. It is describing where most companies land: the tool that works in the demo never becomes the workflow the company runs on. Architecture is a plain word for what closes that distance: the workflows redesigned around the tool, the governance that lets people use it with confidence, and the adoption work that brings the team along. None of it is the model. All of it is the organization.

This is one of the most consistent findings in the research. When McKinsey tested more than 30 factors against real business impact from AI, the **intentional redesign of workflows** was one of the strongest contributors of all, ahead of the model chosen and the size of the budget. On the people side, the same pattern holds: what separates the companies that win through their people is management practices, systems, and workflows, not training spend alone. Both towers answer to the same lever.

Here is where the two towers stop being a metaphor. Add the people tower to the technology tower and the structurally supported value for the same 100-person company reaches about **\$1.06 million a year**. The towers carry the load together: AI's largest measured gains flow to developing talent, and engaged teams led by capable managers realize AI value faster.

FIGURE 5
Two towers do not make a bridge



Build one tower and the value has nowhere to cross. The span is not the model — it is the organization.

Structurally supported values, 100-person company. Source: the Living ROI Calculator, changeadvisor.net.

Build only the technology tower and you are chasing a people-dependent return without investing in the people. Build only the people tower and you leave the largest new productivity lever in a generation on the table. The architecture is the span that lets the value cross.

Governance is the quiet half of it. Only **21% of organizations** planning to deploy AI agents report a mature governance model, even as roughly three-quarters expect to use those agents within about two years. Governance is often what unlocks AI value at the mid-market, because people use a tool they trust and quietly route around one they don't.

21%

MATURE GOVERNANCE

74%

EXPECT AGENTS BY 2027

THE QUIET HALF OF THE ARCHITECTURE

People use a tool they trust and quietly route around one they don't. T2 · DELOITTE, JAN 2026

If this sounds like a large-enterprise problem, the evidence says the opposite. Mid-market firms scale faster, externally partnered builds reach deployment about twice as often as internal-only efforts, and a 100-person company has fewer layers between the decision and the desk. The gap is more closable here, not less.

FIGURE 8
The gap is more closable in the mid-market

DAYS TO SCALE A WORKING PILOT

Top-performing mid-market firms



Large enterprises



Externally partnered builds reach deployment about 67% of the time, against 33% for internal-only efforts.
T2 · correlation, not proof of cause

T2 · MIT NANDA, July 2025. Appendix B, Competitive and Leadership.

What a leadership team does with this

Closing the gap is a short list of decisions, not a transformation program:

-
- 01** **Name the two or three workflows** where AI-addressable work actually concentrates, and redesign those first.

 - 02** **Put light governance in place** so people use the tools without waiting for permission.

 - 03** **Fund the managers** who will carry adoption, because 70% of a team's engagement traces to them.

 - 04** **Measure the gap** — the realized number and the supported number — so the board conversation is about a figure you can defend, not one you hope holds.
-

The distance between what AI could return and what most companies get from it is real, measurable, and not made of technology. A ceiling you have hit is a limit. A gap you can close is a decision. The work starts well before the technology.

The worked example: a 100-person mid-market company

HEADCOUNT	FULLY LOADED	AI INVESTMENT	ADDRESSABLE WORK
100 · 10 managers	\$100K · \$180K	~\$100K / yr	60% × 35%

LINE	AS REALIZED	STRUCTURALLY SUPPORTED
■ Technology tower	\$157,500	\$315,000
■ Engagement gain	—	\$180,000
■ Retention	—	\$375,000
■ Leadership coaching	—	\$172,800
■ Financial wellbeing DIRECTIONAL	—	\$15,658

ARCHITECTURE GAP · TECHNOLOGY TOWER \$157,500 / yr

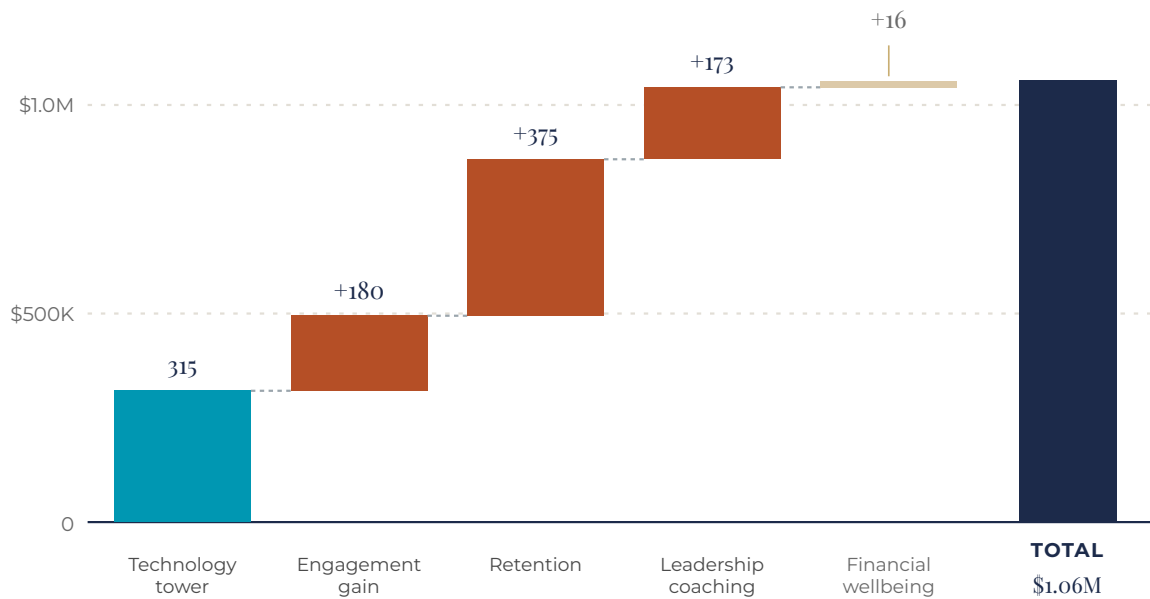
COMBINED, STRUCTURALLY SUPPORTED \$1,058,458

AI PAYBACK	THREE-YEAR AI NET	INVESTMENTS, TRACKED SEPARATELY
~3.8 months	~\$645,000	\$60K · \$25K · ~\$100K

Base 15% uplift scenario · +10 engagement points · 12% stress-cost recovery. Values are gross value produced, with investment costs shown separately rather than netted against the towers. Source: the Living ROI Calculator, changeadvisor.net, verified July 2026.

FIGURE 7

How a 100-person company reaches ~\$1.06M a year



Thousands of dollars per year, gross value produced.

Teal = technology tower · terracotta = people tower · muted = directional.

Bars reconcile to \$1,058,458. Appendix A inputs. Source: the Living ROI Calculator, changeadvisor.net.

ON THE FINANCIAL-WELLBEING BAR

Financial wellbeing is shown at its gross value and flagged directional: no peer-reviewed causal proof exists yet, and net of its \$25K program cost it runs slightly negative. It stays in the chart, labeled, rather than inflated or hidden.

Every bar above is a calculator output at gross value, which is why nothing has to be plugged to reach the total. The investment side — \$60,000 in coaching, \$25,000 in financial wellbeing, about \$100,000 in AI — is tracked separately rather than netted against the towers, so the two views stay legible: what the work produces, and what it costs to run.

The evidence ledger

Every figure grades to a tier. Contested findings are kept and labeled, never dropped. Organized by the five pillars of the AI Transition Maturity Model.

T1 Peer-reviewed experiments **T2** Large-N institutional research **T3** Secondary, flagged

CONTESTED Kept and labeled

TECH / DATA

T2	88% of organizations use AI in at least one function (generative AI about 70%).	McKinsey, Nov 2025 · Stanford AI Index 2026
T2	About 65% have not begun scaling AI enterprise-wide.	McKinsey, Nov 2025
T2	About 95% of enterprise generative-AI pilots show no measurable profit impact.	MIT NANDA, Jul 2025 · methodology
	CONTESTED	

CULTURE

T2	21% report mature governance for AI agents; about 74% expect agent use by 2027.	Deloitte, Jan 2026
T2	20% global employee engagement, a second annual decline.	Gallup SOGW, Apr 2026
T2	About \$10T annual cost of low engagement (roughly 9% of GDP).	Gallup, Apr 2026
T2	Management systems, not training spend, distinguish People+Performance winners.	McKinsey Global Institute, Feb 2023
T1	Generic workplace wellness: no significant causal effect (two RCTs).	Song & Baicker, JAMA 2019
T2	Wellbeing linked to productivity and retention (meta: 339 studies, 1.88M employees).	Krekel, Ward & De Neve, 2019

COMPETITIVE

T2	39% report enterprise-level EBIT impact from AI; about 6% attribute more than 5%.	McKinsey, Nov 2025
T2	Mid-market about 90 days versus enterprise 270+ days to scale a pilot.	MIT NANDA, Jul 2025
T2	\$581.7B global corporate AI investment, 2025.	Stanford HAI, Apr 2026
T3	3% of organizations report a 10–20% AI return; the majority report 1–5%.	Forbes Research, Oct 2025
T2	Top-100 wellbeing firms versus S&P 500: +20% (2021–2023). CONTESTED	Oxford WRC WP2304, 2023 · predictive, not causal

APPENDIX B, CONTINUED

WORKFORCE

T1	Developer speed about +56% (RCT, task-level).	arXiv 2302.06590, 2023
T1	+26% completed tasks, AI coding assistant (3 field RCTs, n≈4,867).	Cui, Demirer, Jaffe et al., 2024
T1	+15% support issues per hour (+30% for the least experienced).	Brynjolfsson, Li, Raymond, NBER/QJE 2023–2025
T1	Consultants: quality about 40% higher inside the frontier; 19% less likely correct outside it.	Dell'Acqua et al., HBS 2023
T1	Writing about 40% faster, quality +18%, the weakest gain most.	Noy & Zhang, Science 2023
T1	AI compresses the skill distribution (consistent across studies).	Multiple, 2023–2025
T2	Engagement premium: +18% sales productivity, +23% profitability (top vs bottom quartile).	Gallup, 2024
T2	People+Performance winners: about 5 points lower attrition (n=1,800).	McKinsey Global Institute, Feb 2023
T1	Coaching effect on performance $g=0.60$.	Theeboom et al., 2014
T1	Coaching effect on organizational outcomes $\delta=0.36$ (conservative floor).	Jones, Woods & Guillaume, 2016
T2	59% of employees financially stressed.	PwC, 2026
T3	Employer cost of financial stress \$1,900–\$2,700 per employee. CONTESTED	John Hancock / Manulife · vendor estimate

LEADERSHIP

T2	Externally partnered builds deploy about 67% versus about 33% internal-only.	MIT NANDA, Jul 2025 · correlation
T3	Direct financial impact overtook productivity as the leading AI success metric.	Futurum, 2026
T2	34% deeply transforming with AI; 30% redesigning key processes.	Deloitte, Jan 2026
T2	Intentional workflow redesign: among the strongest drivers of AI impact.	McKinsey, Nov 2025
T2	70% of team-engagement variance attributable to the manager.	Gallup, updated 2026
T1	Leadership training effect on results $\delta=0.72$ (335 samples).	Lacerenza et al., 2017
T3	Middle-manager practices linked to 5-year TSR 3–21×. CONTESTED	McKinsey, 2023 · correlational
T2	75% of HR leaders say managers are overwhelmed.	Gartner, 2024

Full source links live in the calculator's ledger at changeadvisor.net and are refreshed quarterly. Next refresh: September 1, 2026.

A ceiling you have hit is a limit. A gap you can close is a decision.

The Living ROI Calculator

CHANGEADVISOR.NET

Every figure in this paper is cited and refreshed quarterly. Put in your own headcount, salaries, and AI spend, and the model returns the same two numbers it returned here: what you are realizing, and what the evidence supports. Nothing is stored on the server.