

STATE OF THE INDUSTRY

UPDATED JUNE 2026

The State of AI at Work

95%

of corporate AI transformations returned zero measurable impact on profit and loss. This report is about the few that did not.

The 95% — and the playbook of the few that don't fail.

A report you can forward. The numbers are sourced; the method is repeatable.

BLANKCOLLAR.UNIVERSITY
THE EXECUTION WING

What follows is a diagnosis, not a sales deck.

Three findings, one cause, and a method. The figures come from the field — MIT, McKinsey, the Cursor and IBM disclosures, and the three transformations everyone has read about. We report them straight, then show what separated the 5% from the rest.

01	The most expensive number in business Why 95% of transformations returned nothing	p. 03
02	Everyone is in the gym; almost no one is stronger 78% adoption, 80%+ no earnings impact	p. 05
03	Three bodies on the table McDonald's, Air Canada, Klarna — three different organs	p. 06
04	The real cause: a learning gap, not a weak model What IBM's HR automation actually proved	p. 08
05	The Stack — and the layer you ignored Five layers, one drawing, one exponent	p. 09
06	The economics: work about work, and the leverage gap 58% lost, and \$3–7M per head	p. 11
07	The six moves of the 5%, and the 90-day protocol Buy, aim, clean, keep, redeploy, cancel	p. 14
08	Score yourself, and what the company becomes The four bands, out of 200	p. 16

In 2025, someone finally counted. The result is the most useful fact in business today.

For three years the industry ran on faith. Budgets were approved on a slide and a feeling: that intelligence, dropped into a company like a new hire, would pay for itself. Then MIT's NANDA team did the thing the sellers had avoided. They measured the profit-and-loss impact of corporate AI transformations, at scale, across the economy. The number they found should be printed on the wall of every boardroom that has signed a transformation cheque.

Ninety-five per cent of those transformations returned **zero measurable impact on profit and loss**. Not a soft return. Not a return that arrived late. Nothing a CFO could find. The failures did not come from critics with an axe to grind – they came from the same consultancies selling the cure. And critically, they did not trace to weak models or a shortage of talent. They traced to a **learning gap**: the tools could never absorb how the company actually works.

A 95% failure rate is not a performance problem you fix with another workstream. It is a mortality rate. This report takes that number seriously – and then asks the only question that matters. What did the 5% do differently, and can it be copied?

THE THESIS

Most AI transformations should be cancelled, not improved.

95%

returned zero P&L impact. MIT NANDA, 2025.

5%

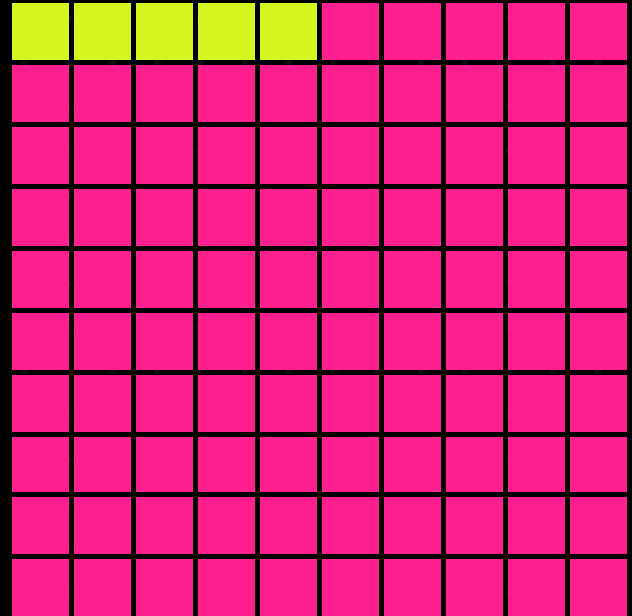
did something specific, and repeatable. The rest of this report is them.

AI does not subtract. It multiplies whatever you point it at.

THE BLANK COLLAR · 2ND EDITION

Out of a hundred, **five** moved the needle.

The waffle is not a metaphor. Ninety-five squares are real transformations that returned nothing their finance team could measure. Five are the ones that did. The gap between them is not budget, and it is not model quality. It is method.



Share of corporate AI transformations by measurable P&L impact. Source: MIT NANDA, 2025.

■ Zero P&L impact — 95

■ Measurable return — 5

The honest reading is uncomfortable: if you are running a transformation right now and have not isolated a single P&L line it moves, the base rate says you are a pink square. The good news is that the five green squares did not get lucky. They followed an order. We can read it off them.

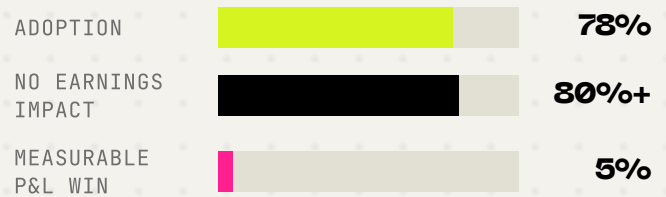
Everyone is in the gym. Almost no one is getting stronger.

The strangest thing about the failure is how busy it looks. Adoption is not the problem – adoption is sprinting. By McKinsey’s own survey, more than **78% of companies** now use AI in at least one function. Tools are everywhere; pilots are everywhere; the dashboards are green.

And yet, in the same survey, **over 80% report no material impact on earnings**. Two numbers that should move together have come apart. Usage is near-universal; results are rare. The activity is real and the outcome is absent, which is the exact signature of motion without leverage.

This is the trap the report keeps returning to. A company can adopt AI completely and change nothing, because it added a powerful tool on top of a process the tool could not understand. Effort went up. Output per person did not. **Being in the gym is not the same as getting stronger.**

THE TWO CURVES THAT CAME APART



Adoption and earnings impact: McKinsey, 2025. P&L win: MIT NANDA, 2025.

WHAT IT MEANS

If adoption alone worked, 78% would be winning. The result lives in the method, not the tool.

A powerful exponent on a base below one does exactly what exponents do. It makes small things smaller, faster.

ON THE 95%

Three celebrated projects. Three different organs. Not one killed by the AI itself.

The failures are usually filed under “the technology was not ready.” The autopsies say otherwise. In each case the model did precisely what it was built to do. What broke was a layer that was rotten before the technology arrived.

McDonald's x IBM

Put an order-taking model on a drive-thru built on a process no one had cleaned. You cannot automate a mess – the mess wins. Bacon on the ice cream, and the pilot was pulled. The fault was upstream of the AI.

PROCESS FAILURE

Air Canada

A support chatbot invented a refund policy that did not exist, and the airline tried to disown what its own agent had said. A tribunal disagreed. You own what your agents say – governance is not optional once the machine speaks for the brand.

GOVERNANCE / HX

Klarna

Replaced human support to chase the cheapest possible answer, declared victory, then quietly began rehiring as quality and satisfaction fell. The cheapest answer turned out to be the most expensive one. The failure was in the aim, not the engine.

VISION FAILURE

THE PATTERN UNDER ALL THREE

Point a powerful model at a clean foundation and it compounds. Point it at a broken one and it industrialises the break. The technology was never the variable. The layer beneath it was.

THE COUNTER-CASE

IBM automated roughly **94% of routine HR tasks** – and total employment **rose**. The hours were redeployed, not deleted. Aimed well, AI does not subtract.

02

The real cause is a learning gap, not a weak model.

If the models worked and the companies still failed, the fault is not in the technology. It is in the thing the technology was asked to multiply. To see it, you need one drawing – and one piece of arithmetic about exponents.

The specialist is now a training set. The generalist who directs the machine is the asset.

Specialisation was never a law of work. It was an adaptation to one condition – that knowledge was expensive to acquire and impossible to copy. That condition has collapsed. A narrow, deep, perfectly repeatable expertise is no longer a fortress in the AI era; it is a training set. The more legible and rule-bound your work, the more completely a model absorbs it. Obscurity used to be job security. Now it is how you get automated first.

This is why the failures were a learning problem. The companies bought capability and pointed it at processes that lived in people's heads, on data that contradicted itself, in service of a vision no machine could read. The tool could not learn the company, so it amplified the confusion at speed.

Recall is dead; judgment is orphaned. The value was never the memorised table – it was knowing when to use it.

The winners drew the opposite conclusion. They stopped hiring people to perform tasks and started hiring people to design the systems that perform them. The role defined by doing the work gets cheaper as machines get cheaper. The role defined by deciding what work happens, and aiming the machines at it, gets **more valuable with every price cut.**

THE IBM DISCLOSURE, READ CORRECTLY

94%

OF ROUTINE
HR TASKS
AUTOMATED



TOTAL
EMPLOYMENT
ROSE

IBM workforce disclosure. The hours were redeployed, not eliminated.

THE ORCHESTRATOR'S LOOP

Specify the outcome ·
compose the agents and tools
· verify against a broad sense
of good. Then design yourself
out of the seat.

The losers ask how to add AI. The winners ask what they would build if labour cost almost nothing.

THE TWO QUESTIONS

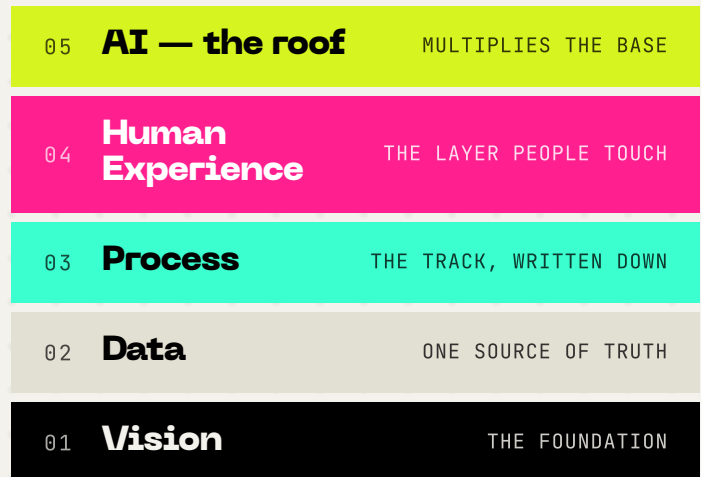
The Stack. You break at the layer you ignored.

Flat frameworks fail because they are lists posing as diagrams — five equal boxes that let a company say “we’re strong on three of five, so we’re 60% there.” Real systems are topologies, where one thing must be true before the next can be. The Stack is five layers you can draw on a napkin, and the order is the whole argument.

Vision is the foundation: one sentence sharp enough that a person — or an agent — can reject a bad project without asking. **Data and Process** are the load-bearing middle: the fuel, and the track the work runs on. **Human Experience** is the layer your people actually touch. And **AI is the roof** — not a fifth box, but the layer that multiplies whatever is underneath it.

That placement is the diagnosis in one stroke of a pen. Almost every failed transformation broke at exactly one layer, and it was always the unglamorous one the company skipped. The weak layer is never a secret to the people doing the work — **only to the leadership funding the layers it found interesting.**

BOTTOM-UP: STABILISE BEFORE YOU TURN ON THE ROOF



THE RULE

A perfect roof on no foundation is 0% there, not 60%. The lowest layer caps everything above it.

Stabilise in order: Vision first, then Data and Process together, then the human layer. Only with a base genuinely greater than one do you turn on the multiplier. The 95% reversed it — they bought the roof and left the foundation cracked.

Same exponent. Opposite outcome. The base decides.

AI is not an ingredient you add. It is an exponent, and it is indifferent to your intentions — it multiplies whatever base you hand it, and carries the base's sign.

BASE ABOVE ONE

$$2^{10} = 1,024$$

A clean foundation, raised to the tenth power. Small advantages compound into a company that pulls away.

BASE BELOW ONE

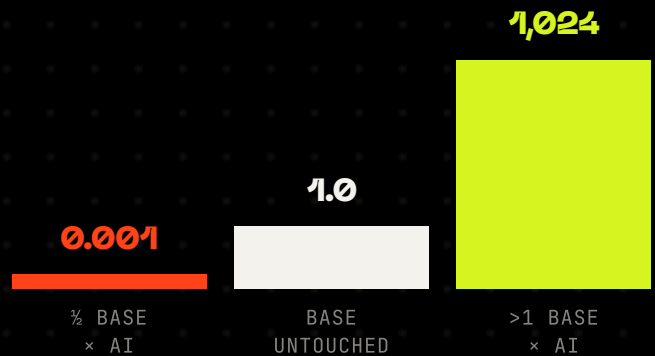
$$1/2^{10} \approx 0.0001$$

A broken foundation, same exponent, drops to almost nothing. The roof does not rescue the house; it schedules the collapse.

This is the precise reason the 95% failed. They did not buy bad AI. They applied a powerful exponent to a base below one — and the exponent did exactly what exponents do. **It made small things smaller, faster.**

Point a capable model at a messy process and you do not get a fixed process. You get a faster, more confident mess — fluent, well-formatted, and wrong at scale. The model is faithful. That is the danger.

WHY “WE NEED BETTER AI” IS THE WRONG RESPONSE



Illustrative: the same multiplier on three different bases.

A stronger exponent on a base below one produces a worse result, faster.

THE LAW OF THE ROOF

You pay for a full week of expert labour and get back about two days of it.

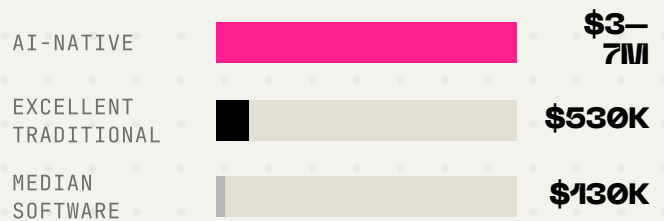
58%

of the average knowledge worker's day goes to **work about work** – searching, chasing, and confirming which number is real. The tax is invisible on every statement; it hides inside salaries and the vague sense that everything takes longer than it should.

When there is no single source of truth, there is no truth – only negotiation, and the most senior person's spreadsheet wins. Three tribes keep three sets of numbers, none convertible. This is the Data layer failing quietly – the most expensive line item no accountant records.

Knowledge-worker time on “work about work”: industry research cited in The Blank Collar.

REVENUE PER EMPLOYEE – THE NEW SCOREBOARD



Revenue per employee: AI-native (Cursor) vs. an excellent traditional software firm vs. the median. Not harder work – more leverage.

READ THE RATIO, NOT THE REVENUE

A company that doubles revenue by doubling staff has grown, not scaled – and quietly doubled its coordination cost. One compounds; the other inflates.

Headcount stopped being a measure of strength. Carried past what the work needs, it is drag.

ON SCALING WITHOUT WEIGHT

The jagged frontier: the same task, a step apart, and the result inverts.

In a controlled study, consultants given AI on tasks inside the model's range pulled away. On a task that looked almost identical but sat just outside it, they did worse than the colleagues working without it – because the model produced confident, plausible, wrong answers, and the humans trusted them.

INSIDE THE FRONTIER

More tasks completed	+12%
Faster	+25%
Higher quality	+40%

Weakest performers gained the most

OUTSIDE THE FRONTIER

Accuracy vs. no AI	worse
Failure mode	confident
Visible from outside	no

The failures look exactly like the successes – until a human checks

The lesson is not “use less AI.” It is that the human does not leave the loop – **the human moves up it**. Inside the frontier, let the machine carry the load. At the edge, where confident wrongness lives, judgment becomes the entire job.

This is why the order in the Stack matters so much. A clean base keeps work inside the frontier; a messy one pushes it past the edge, where the model is most fluent and least correct. **Reliability is a discipline, not a model upgrade.**

03

The playbook of the few that don't fail.

The 5% are not smarter or richer. They are the companies that did not make the three mistakes – and they ran a repeatable sequence you can start this quarter. Six moves, ninety days, four bands.

What the 5% did — and the rest skipped.

The winners are simply the companies that avoided the three deaths and applied the exponent last. None of these moves needs a bigger model. Every one of them is a decision, available today.

01 Buy before you build.

Bought tools reach production about twice as often as in-house science projects. Buy the capability that works; do not fund a research department by accident.

02 Aim at one P&L number a CFO can find.

Not “a transformation.” A single line, with a date. If you cannot name the number it moves, you do not have a project — you have a budget line for theatre.

03 Clean the process before you automate it.

Write it down; run the five-whys; delete the ghost steps. You cannot automate a mess, a vibe, or a habit with good PR — only what you have made explicit.

04 Keep a human where the brand lives.

Air Canada owned what its agent said the hard way. Accountability cannot be delegated to a tool. Put a person on the moment that carries the brand.

05 Redeploy the savings into work that compounds.

IBM's headcount rose because the hours moved up the value chain. Point the freed time at judgment, novelty, and trust — not at a smaller payroll.

06 Cancel fast.

A dead initiative is hygiene, not shame. Any pilot that cannot survive moves one to five is already finished — sign the certificate and free the budget for the one that can.

Nobody is born into the five percent. You choose your way in.

THE BLANK COLLAR

A method should end with a move. Here is the ninety days.

No competent doctor prescribes before they diagnose. The protocol finds the truth, raises the base above one, then turns on the multiplier – in that order, every quarter, forever.

DAYS 1-30

Diagnose

Score your Stack and name the red layer. Cancel any initiative with no P&L line. Compute revenue per employee and write it next to the benchmarks.

FIND THE WEAKEST LAYER

DAYS 31-60

Fix the base

Run the red layer's 30-day move. Encode one process. Free your three data hostages. Run one two-week agent pilot from where you sit, inside policy.

RAISE THE BASE ABOVE ONE

DAYS 61-90

Compound

Redeploy what you saved. Hire one orchestrator, not an operator. Re-score the red layer. If it moved, you have proof – and a method you can run again.

TURN ON THE MULTIPLIER

Grade the audit honestly – true on a bad week, not just on board day. Then read the shape, not the total. A company scoring 38 on AI and 12 on Data does not have a strong AI capability. **It has a fast hallucination machine, and its real score is the 12.**

THE THREE QUESTIONS TO ANSWER BY QUARTER-END

What is our revenue per employee, and the target that means we need not hire? What is our red layer, and the one 30-day move? What would we build if labour cost almost nothing – and what is stopping us this quarter?

Score the five layers. Read the band. It is your roadmap.

Forty statements, eight per layer, each scored one to five. Sum each layer, then find the lowest – that collapsed point caps everything above it. Out of 200, four bands tell you the truth.

40–99**The 95%**

Zero on the P&L. Stop funding the roof; repair the foundation. A stronger model here only speeds the loss.

**100–
139****Fragile**

A base that holds on a good week and slips on a bad one. One shock from sliding back. Stabilise before you scale.

**140–
169****A contender**

A real base above one. Find and fix the single layer dragging the rest, and the exponent starts to pay.

**170–
200****The 5%**

Compounding. The roof is multiplying a base worth multiplying. Your only real risk now is complacency.

The Index is the quick, twenty-question version of this audit – a measured read, scored for you, in about five minutes. It places you on the band and names the layer to fix first.

RUN THE
INDEX ·
FREE

When the base is real, the company changes shape.

Past the 90 days, the firms that compound start to look different from the ones that did not. They scale without weight: output per person rises while headcount stays flat, because the leverage lives in the architecture, not the org chart. Every hire is treated like a kilogram of payload — added only when no system can do the job.

They own their intelligence. Renting the frontier is fine for the rare hard problem; surrendering your core cognition to a third-party API is building on rented land. Open-weight models now run within months of the closed frontier, and for most business functions the difference is immaterial — so the routine many run on models they control, and only the rare hard problem is rented.

And they redraw the work around the human 1%. As the machine takes the routine 99%, the final 1% — the judgment, the relationship, the decision about what is worth building — is not the scraps. **It is everything that ever mattered, and it expands as the machines take the rest.**

WORK IS FOR BOTS

- The routine 99% — drafting, reconciling, classifying, routing
- Process that fires automatically, in code, not in decks
- The bridges between systems no person should carry

LIFE IS FOR HUMANS

- The high-stakes judgment at the edge of the frontier
- The conversation where someone must be trusted
- Deciding what is worth building at all

The operating principle, not the slogan: the human keeps the 1% where the value always lived.

Work is for bots. Life is for humans — not as a slogan, but as an operating principle.

THE BLANK COLLAR

BLANKCOLLAR.UNIVERSITY - THE EXECUTION WING

You now know the base rate. The method is the part that changes it.

The school is the execution wing of The Blank Collar: the programs, formats, rubrics and outcomes that turn this report into a base above one. Diagnose your Stack, fix the red layer, and run the ninety days with people who have done it.

[RUN THE EQUATION · THE-INDEX](#)[BOOK A TRAINING · /BOOK](#)

**Nobody is born into the five percent.
You choose your way in.**

BLANKCOLLAR.UNIVERSITY
THEBLANKCOLLAR.COM