BURAK GALİBA
Thoughts on AI Transformation

ESSAY · ENTERPRISE AI

AI ROI — If You Didn't Baseline It, It Didn't Happen

95% of AI pilots can't prove ROI, mostly because nobody measured anything first. A simple baseline discipline that makes AI value undeniable.

ROI AI STRATEGY

Burak Galiba · June 16, 2026 · 4 min read
https://burakgaliba.com/blog/how-to-measure-ai-roi

SCAN TO READ ONLINE

Ask an executive team whether their AI program is working and you'll hear real enthusiasm: the team loves it, adoption is strong, the demos keep getting better.

None of that is a financial metric. And when the CFO finally asks the financial question — increasingly with $2.5 trillion of worldwide AI spend demanding justification — the room discovers that nobody can answer it. Not because the AI failed, but because nobody measured anything before it started.

That's the quiet scandal inside the widely cited MIT NANDA finding that 95% of enterprise GenAI pilots show no measurable P&L return. "No measurable return" has two failure modes: the value wasn't there, or the value was never measured. Both end the same way at budget time.

The baseline is the cheapest thing you'll skip

ROI is arithmetic on two numbers: the world before, and the world after. Skip the "before" and no amount of after-the-fact analysis recovers it. You'll be estimating your own past from memory, and finance will treat those estimates exactly as they deserve.

The before-state costs almost nothing to capture — it's usually a week of pulling numbers you already have:

- **Cycle time.** How long does an invoice, claim, ticket, or contract take today, wall-clock, end to end?
- **Cost per transaction.** Fully loaded — people, systems, rework.
- **Error and rework rates.** What fraction comes back, gets escalated, or gets corrected downstream?
- **Volume per person-hour.** Throughput of the team as currently staffed.

Notice these are workflow numbers, not AI numbers. That's the point: AI ROI is always measured in the currency of the process it touches. If a proposed use case can't name which of these numbers it will move, it isn't a use case yet — it's a technology looking for a chaperone.

Count all the costs, not just the invoice

The other half of honest ROI is the denominator, and token bills are the *small* part of it. A credible cost line includes:

- Integration engineering — usually the largest single item
- Evaluation and testing before launch, and monitoring after
- Human review time for everything the AI escalates or gets wrong
- Change management: training, workflow redesign, the productivity dip during transition
- Vendor and platform fees, including the ones bundled into cloud commitments so they feel free

Enterprises that count only the license fee report thrilling ROI right up until finance rebuilds the number. Do the full accounting yourself, first — an honest 2x beats an imaginary 10x in every meeting that matters.

Give the number an owner and a deadline

A measured metric with no owner is trivia. The discipline that actually changes outcomes is a 90-day loop:

1. **One metric per use case.** Not a dashboard of eleven — one number a named executive already cares about.
2. **Baseline before build.** No exceptions, no retroactive reconstructions.
3. **Instrument the workflow**, so the metric updates from live data rather than quarterly anecdote.
4. **Review at day 90 with three verdicts** — ship it wider, fix the named blocker, or kill it and reclaim the budget.

This is also how the market now behaves. Enterprise AI procurement has professionalized: outcome gates, governance requirements, business cases treated like core software purchases. Even the consulting giants have conceded the point — McKinsey now ties 25–30% of its global fees to outcomes rather than hours. When the sellers of advice accept payment on results, buyers who can't measure results are negotiating blind.

The strategic payoff of boring measurement

Here's what the baseline discipline actually buys you, beyond CFO peace.

It changes what gets funded — demo-charisma projects lose to measurable ones, which quietly steers your portfolio toward the back-office work where that study found the real returns. It makes kill decisions unemotional, because the number decides instead of the sponsor. And it compounds: the first proven use case makes the second one's business case a formality.

Enterprises with baselines argue about expansion. Enterprises without them argue about belief. Only one of those arguments survives a budget cycle.

Where to start

Don't build anything this week. Pick the one AI initiative you'd most like to defend at year-end, and reconstruct its baseline now while the pre-AI data still exists — cycle time, cost per transaction, error rate. It's a few days of work that converts your best project from a story into a number.

For a broader read on whether your organization measures like the 5% or hopes like the 95%, my free AI Readiness Score covers exactly this — 20 questions, about ten minutes, and a scored breakdown that shows exactly where your measurement discipline stands today.

Read the essay online at https://burakgaliba.com/blog/how-to-measure-ai-roi

I built a free AI Readiness Score — 20 questions across pilots, data, talent and governance, about ten minutes, no email required: https://burakgaliba.com/readiness