



PAPER • KNOW WHAT IT RETURNED

Measuring AI in dollars

How to tell whether an AI system paid, from the baseline you take first to the ledger that runs after launch.

StrikeArc Labs · October 2026

strikearc.ai/resources/measuring-ai-in-dollars

Measuring AI in dollars: how to tell whether it paid.

Every vendor promises a return on AI. This paper shows how to check one, from the baseline you take before anything changes to the ledger that keeps counting after launch. It is written for an owner, and for the CFO who will be asked to sign off.

STRIKEARC LABS · OCTOBER 2026 · 14 MIN READ **The paper** PDF

Most owners have been promised a return on AI by now. Far fewer have seen one proved. When a return cannot be proved, the cause usually sits in how the numbers were made: nobody measured the work before it changed, or nobody would put their name to the figure afterward.

This paper sets out a standard for measuring an AI system in dollars. Use it to judge any proposal on your desk, ours included. Each rule is a habit you can write into a contract or a project plan, and none of them needs special software.

THE STANDARD IN BRIEF

1. Take the baseline before anything changes, and mark how sure you are of each number.
2. End every candidate workflow with a verdict, and let do-not-build-yet be one of them, with the condition that would reopen it.
3. Where an input is uncertain, show a low and a high figure.
4. Trace every figure to an input you gave or a record you can open.
5. Put a named person's signature on every number before it reaches a decision.
6. Carry the promise into the install: the baseline, the target, what counts as done and when the first proof is due.
7. Keep a ledger after launch, and treat a system that stops reporting as a warning.

Why AI returns are hard to believe.

Four habits account for most of the returns that cannot be checked. Each is common. Each is cheap to fix before the work starts, and close to impossible to fix after it.

01 **Nobody measured the work before it changed.**

The process that existed before the install is gone the day the install goes live. If no one timed it, counted its errors or wrote down how long a customer waited, every later figure for what was saved is an estimate built on a memory. Some providers fill the gap with an industry benchmark. A benchmark describes someone else's business.

02 **Activity gets counted as a result.**

Messages sent, documents processed, tickets touched, calls logged. These counts rise the day a system starts, whether or not anything got cheaper, faster or more profitable. They prove the system ran. A result needs a second step: what each of those documents or messages was worth against the baseline, in dollars.

03 **Nobody signs the figure.**

A number on a dashboard with no owner can be wrong for months. A formula breaks or a definition changes, and nobody notices. When a figure carries a person's name, someone has checked where it came from and will answer for it when your CFO asks.

04 **The cost to run it is left out.**

An AI system keeps costing money after launch: usage fees for the models, and the hours someone spends checking its work and fixing it when an upstream system changes. A return quoted before those costs is a gross figure. The number to decide on is what is left once they are paid.

The standard below answers each habit with a rule. Most of the rules cost a few hours at the start of a project. Skipping them costs you the ability to prove anything at the end.

The standard, in seven rules.

Hold any provider to these, and hold your own team to them when you build in-house. They follow the order of a project: before the work, at the decision, during the install and after launch.

01 Take the baseline before anything changes.

For each workflow in scope, write down who does the work, how many hours a week it takes and what those hours cost in a year at the loaded rate. Next to every number, mark how sure you are of it: measured, estimated or a guess. The baseline is the line every later claim is measured against, so it has to exist before the first change, written down where both sides can see it.

02 End every candidate workflow with a verdict.

Each workflow you consider ends in one of three places: build it, build part of it, or do not build it yet. The third verdict is the test of an honest assessment. It should appear even when it costs the provider the project, and it should name the condition that would reopen it, such as an ERP migration finishing or an owner being named.

03 Show a low and a high figure where an input is uncertain.

A single number to the dollar looks precise even when the inputs behind it are guesses. Where an input is estimated or guessed, give the return as a low and a high figure, and say which input moves it most. A wide range tells you what to measure next. It is also much harder to oversell.

04 Trace every figure to something you can check.

Every dollar in a readout should lead back to an answer you gave, or to a record from your own systems for a named month. A figure that traces to nothing gets left out. A readout built this way is often shorter, and it holds up when your CFO starts asking where each line came from.

05 Put a named person's signature on every number.

Before a figure reaches a decision, a person checks its source and its arithmetic and signs it. The signature tells the reader who to ask, and it makes the person signing look twice. The same principle covers what a system does: anything that moves money or reaches a customer waits for a named approver.

06 **Carry the promise into the install.**

The workflow chosen to go first should leave the readout with its promise written down: the measured baseline, the target after the build, the return expected in a year, what counts as done and the date the first proof is due. That record becomes the test the install has to pass, so what was promised is what gets measured.

07 **Keep a ledger after launch, and listen for silence.**

A live system reports on a schedule: what it handled, what it cost to run and whether its work still meets the bar set at the install. Each period, the ledger turns that into dollars, lever by lever, with the cost to run on its own line and a person's name on the result. A system that stops reporting has stopped working or stopped being watched, and either one should raise an alert before anyone reads a quality score.

What a baseline looks like.

A baseline is a short table, one row per workflow. Here is one row for vendor invoices at a made-up distributor, with the same inputs as the worked example in the [AI readiness assessment guide](#). The column most baselines leave out is the one that says how sure you are.

INPUT	VALUE	HOW SURE	WHERE IT CAME FROM
Invoices a month	2,000	Measured	The accounting system's count for last month
Minutes of AP time each	7	Measured	A week of invoices timed by the AP lead
Loaded cost of an AP hour	\$38	Estimated	Salary from payroll, plus benefits, payroll taxes and overhead
AP time a month	233 hrs	Measured	Invoices times minutes, divided by 60
What the work costs a year	\$106,400	Estimated	AP time times the loaded cost, times 12

Every figure here is invented to show the method. Yours come from your own records.

Two habits make the table useful. A number worked out from other numbers is only as sure as its weakest input, so the yearly cost above is an estimate even though most of its inputs were measured. And the guesses get checked first: if the loaded rate is a guess, an hour with the payroll records turns it into an estimate before any decision rests on it. How to build a loaded rate, and the rest of what an invoice costs, is in [how to calculate the cost to process an invoice](#).

If you are still choosing which workflows deserve a row, start with the five tests in [which business processes to automate first](#), or score them in [the Workflow Shortlist](#).

What a promise looks like when it is carried forward.

Say the readout recommends starting with those invoices. The promise that leaves the readout is a short record, and it becomes the acceptance test for the install. Continuing the made-up example:

Workflow	Vendor invoices: read, matched to the purchase order and the receipt, and posted
Baseline	233 hours of AP time a month, measured
Target after the build	80 to 110 hours a month, with every exception routed to its owner
Expected return in a year	\$56,100 to \$69,800, before the cost to run it
Where the hours go	To the work the AP team has waiting, named by the AP lead before the build
What counts as done	Routine invoices post without a person touching them. Every exception reaches its owner with the evidence attached. Nothing posts outside the approval rules.
First proof due	The first month-end close after go-live

Every figure here is invented to show the method. The range comes from the two ends of the target: 123 to 153 hours a month handed back, at \$38 an hour, for twelve months. They count only because the record says where they go.

When the install finishes, this record is the checklist. If AP time lands at 140 hours a month, everyone can see the shortfall, and the conversation is about why. Nobody has to reconstruct what was promised from a slide that has since been edited.

What each lever is counted in.

Every return lands on one of three levers, and each lever is counted in its own way. Counting the same dollar on two levers is the most common way a ledger inflates.

LEVER	WHAT MOVES	HOW IT BECOMES DOLLARS
↑ GROW REVENUE	Revenue that stops slipping away: inquiries answered before a competitor calls back, quotes that go out the same day	Revenue won or kept, times your margin on it
↓ REDUCE COSTS	Money that stops leaking: duplicate payments, billing errors, unclaimed deductions, AI spend nobody uses	Dollars recovered, or no longer spent
↓ SAVE TIME	Hours handed back to the people who do the work today	Hours times the loaded cost of an hour, where the hours go to other work

Saved hours need the most care. An hour is worth its loaded cost only when it goes somewhere: to work that was waiting, or to overtime and hires you no longer need. Write down where the hours go when you count them. If they go nowhere, count them at zero and say so.

Each play in the StrikeArc PowerPlays Library names what it reports. Some of those lines are dollars already. Others are operating measures, such as the share of documents processed without a touch, and the ledger turns them into dollars through the baseline: a document that no longer needs a person is worth the minutes it used to take, at the loaded rate, counted once.

PLAY	LEVER	WHAT IT REPORTS
P01 Put your ad budget behind what works	↑ GROW REVENUE	Share of spend on proven messages, and what it did to acquisition cost.
P02 Prepare business reviews in less time	↓ SAVE TIME	Hours returned to the exec team, and forecast accuracy against actual.
P03 Win more deals with better design	↑ GROW REVENUE	Win rate on the redesigned surfaces.
P04 Stop overpaying for AI	↓ REDUCE COSTS	Dollars of AI spend recovered.

P05 Find the money you're leaking

↓ REDUCE COSTS

Recovered dollars.

P06 Cut the manual work from orders and invoices

↓ SAVE TIME

Share of documents processed without a touch.

P07 Respond to new leads faster

↑ GROW REVENUE

Answer time, booked jobs and revenue per lead.

P08 Get quotes and contracts out faster

↑ GROW REVENUE

Quote turnaround and win rate.

P09 Track the competitor moves that matter

↑ GROW REVENUE

Battlecard freshness and competitive losses.

P10 Spot churn, margin, and cash risks sooner

↓ REDUCE COSTS

Early catches on churn, margin and cash.

P11 Turn your expertise into a product

↑ GROW REVENUE

Expert hours productized and twin usage.

P12 Find the answers your team needs

↓ SAVE TIME

Answer rate, and how fast new hires get productive.

A quarter of the ledger, worked through.

Here is what one quarter of a ledger can look like for a company running three plays from the library. The figures are the illustrative ones on our [How It Works](#) page.

P06 Cut the manual work from orders and invoices: 1,840 documents processed without a touch	\$38,400
P05 Find the money you're leaking: duplicate payments and missed credits recovered	\$41,200
P08 Get quotes and contracts out faster: revenue won on same-day quotes	\$96,000
Value this quarter, before the cost to run	\$175,600

Every figure here is invented to show the method. Your numbers come from your own baseline.

Each line has to trace to something you can open.

LINE TRACES TO

- P06** The documents that posted this quarter without a person touching them, the minutes each one took in the baseline, and the loaded rate
- P05** Each recovered payment or credit, matched to the vendor's confirmation or the bank record
- P08** The quotes sent the same day, the ones the customer accepted, and your margin on them

Revenue lines need the strictest test, because some of those wins would have come anyway. One fair method: compare the win rate on quotes sent the same day with the win rate in the baseline, and count only the difference, times the margin. Write the method down next to the figure, so the next quarter is counted the same way.

The ledger above stops at value. Before \$175,600 is compared with anything, subtract what the plays cost to run that quarter: the fees and usage, plus the hours your team spent checking the work. The net is the figure a person signs, and the figure your CFO should hold against the baseline.

What to ask any provider.

You can test a proposal against the standard in one meeting. These questions work on a consultant, an agency, a software vendor, an in-house team or us.

- What will you measure before you start, and how sure will we be of each number?
- Which of our workflows would you tell us to leave alone, and what would change your mind?
- Where an input is uncertain, will you show us a low and a high figure?
- Can every number in the readout be traced to something we can open?
- Who signs each figure, by name?
- After launch, what gets reported, how often, and in what unit?
- What happens when the system stops reporting?
- Is the return you quote counted before or after the cost to run it?

The full set of twenty questions, with what a strong answer sounds like and a sheet to score up to three providers side by side, is in [the AI Partner Scorecard](#).

How we apply the standard.

Here is where the standard shows up in how we work, in the words the rest of this site uses.

- **Fit comes before value.** StrikeArc Scout scores every play in the StrikeArc PowerPlays Library against your answers. Scout says which plays match, and it puts no dollar figure on them.
- **A person signs the numbers.** Before the readout, one of our people sizes the value at stake and signs every figure. Nothing reaches you that a person has not checked.
- **The readout names what to skip.** A person walks you through the value at stake, the top plays and the do-not-build list. If no play in the library pays its way, the readout says so, and the readout is yours to keep either way.
- **Approvals stay with people.** Every gate is a named human. Agents draft everything; they approve nothing.
- **Results are reported in dollars.** Every play reports its number in dollars, and a person signs it. Where a play's own measure is an operating one, such as documents processed without a touch or quote turnaround, the ledger turns it into dollars through your baseline, as the section above shows. On the subscription, we report the revenue gains, savings and hours returned to your team, and keep each play's cost down as the models change.

Hold us to the other rules as well. Every question in the section above is fair to ask us on an intro call, and a straight answer to each one is part of what you should expect. More on the route from the first call to a running play is in [how it works](#) and on [StrikeArc Scout](#).

Keep reading.

The AI Partner Scorecard: twenty questions to ask before you sign

CHECKLIST

The Workflow Shortlist: find the workflow to automate first

TOOL

AI readiness assessment: how to find where AI will pay in your business

GUIDE 01

Managed AI services: what it takes to keep AI paying after launch

GUIDE 04

Bring the workflow that costs you the most.

Book a 30-minute intro call at strikearc.ai/book, or write to info@strikearc.ai.

The latest version of this piece lives at strikearc.ai/resources/measuring-ai-in-dollars.