

Every AI Budget Is a Bet

A practical brief for executives deciding what AI is worth paying for

Pull up last month's AI charges. Some are subscriptions. Some are usage bills that arrive on a card statement. Some are buried inside software you already bought.

Now try to say, next to each one, what it bought.

The gap between those two columns is the subject of this brief. Every charge is a bet on a return: more capacity, better work, fewer mistakes, something the company could not do before. That isn't a criticism. Most business spending starts with an expectation rather than proof.

Accenture surveyed 750 executives at companies with more than \$1 billion in revenue. Less than one dollar in five of what they spend running AI could be traced to a measured financial result. Only 35 percent could say what a single business result costs, even for their largest AI use.

The question isn't whether you're betting. It's whether you know what you bet, what you expected back, and what would make you change course.

The first half of this brief is about the return on what you already run. The second is about placing the next bet before you can know exactly how it will pay.

EIGHT SECTIONS · TWO PARTS

WHAT YOU ALREADY SPEND

- | | | |
|--------------|--------------|----------------|
| <div>1</div> | <div>2</div> | <div>3</div> |
| THE BILL | THE UNIT | THE RIGHT TOOL |

THE NEXT BET

- | | | | | |
|-----------------|---------------|--------------|--------------|-----------------|
| <div>4</div> | <div>5</div> | <div>6</div> | <div>7</div> | <div>8</div> |
| BEFORE YOU KNOW | THE FIRST BET | WHEN TO STOP | WHAT EXPIRES | WHERE YOU STAND |
-

What you're getting for what you already spend

1

The bill

Two things can push your AI costs up, and they need different answers.

Use can grow faster than prices fall. The executives in Accenture's survey expect their AI usage to grow 78 percent over the next two years while prices fall about 19 percent. If both forecasts hold, the usage bill still rises by more than 40 percent. One in three of those companies already runs out of its AI usage budget before year-end.

A growing bill may be justified. It may mean valuable work is expanding. But somebody needs to be able to tell productive growth from consumption nobody is watching.

The price list doesn't tell you what the work costs. In September, a leading AI vendor cut the price of its top product by 20 percent. In independent testing, the new version cost about the same per task as the old one, because it produced roughly 1.6 times as much output to get there. Your work will differ. The lesson won't: a lower price per unit doesn't mean a lower cost per finished job.

| **A price cut only helps if somebody is counting what the work costs.**

2

The unit

Seats, logins, and monthly invoices tell you what you bought access to. They don't tell you what the company got back.

Choose a unit the business already understands: one invoice coded correctly, one proposal ready to send, one customer question resolved, one month-end report accepted.

The standard matters. A draft that takes an hour to repair is not the same result as one that is ready to use.

Then ask two questions.

What does one cost? The software and usage charges, the time spent checking and correcting the output, and the work of keeping it running. Record setup costs separately so they don't disappear from the investment decision.

What is one worth? Revenue won. A mistake prevented. Faster service. People freed for the work only they can do.

Hours saved are only worth something if they go somewhere. If nobody can say where the time went, you have measured a possibility, not a return.

Some uses will justify more investment. Others will need a redesign or a decision to stop. For many, the first finding is simply that nobody has checked.

The right tool

Once you know what an acceptable result costs, you can ask whether you're using the right tool to produce it.

Most AI-assisted workflows are a chain of smaller jobs: pulling out information, sorting it into a category, checking a rule, drafting an answer, handling the exception. They don't all need the same horsepower. Some don't need AI at all.

Here is the common setup at a mid-sized company. Whoever built the workflow picked the best, most expensive AI available and used it for every step. It works. Nobody asked whether each step needed it.

Accenture studied more than 9,000 business tasks and found that 90 percent didn't need the most capable, most expensive AI. In the same research, 54 percent of AI requests went to a more expensive tool than the task required.

Paying premium prices for sorting work is a design choice. It's usually one nobody knew they made.

That's a reason to test cheaper options, not to downgrade everything. A cheaper tool that creates more checking work can cost more overall.

Choose the least expensive approach that reliably meets the business standard.

BETWEEN THE TWO HALVES

You can't judge a new bet if you can't measure the ones you've already placed.

PART TWO

Placing the next bet

Before you know

You can estimate what an AI project will cost and return. You can't confirm the estimate until it runs on real work, under the conditions your people actually face.

That uncertainty tempts companies into two mistakes: waiting for a promise nobody can responsibly make, or approving a project on enthusiasm and hoping.

Waiting is also a bet. The unsolved problem keeps costing time or money, and the company delays learning what might work.

You don't need to measure everything you already run before testing something new. You need a baseline for the work in question. That starts with one page, written before you commit to the pilot.

The example is illustrative. Its numbers are assumptions to test, not estimates for your company.

WRITE DOWN

ILLUSTRATIVE INVOICE-PROCESSING PILOT

What happens today	About 4,000 invoices a month at a fully loaded cost of \$8 each, 99 percent accurate before final approval.
What an acceptable result is	An invoice coded and ready for approval, at least as accurate as today.
What we expect it to cost	\$2 to \$6 per accepted invoice, including software, checking, corrections, and routine support.
What improvement is worth pursuing	At \$4 an invoice, about \$16,000 a month of the team's time returned, used to clear the backlog of disputed invoices.
What we'll spend to find out	Up to \$6,000 for setup and a 500-invoice pilot over 60 days.
What it makes possible next	Purchase orders, if the invoice pilot meets its targets.
What it commits us to	Daily exception review by finance, ongoing support, and exposure to the vendor's pricing.

The time returned is not a promise of cash savings. The return is what the team does with the time, so name that in advance.

Some returns will stay uncertain. Write down how you'll judge them rather than forcing a number you can't defend. What isn't fine is leaving the page blank.

If you never wrote down what you expected, you'll be tempted to call whatever happened a success.

5

The first bet

The project with the biggest promised payoff isn't automatically the best place to start. Big-payoff projects tend to be the most visible, the most political, and the hardest to measure.

For a first project, look for a bounded test of an important uncertainty. Can it meet the quality standard? Is the checking work manageable? Will people use it in the middle of their actual work? Choose a test small enough to stop, but representative enough to teach you something.

The best first project is the one that tells you the most about whether the second project is worth doing.

That second project may create a capability the company doesn't have today. A team that reads a sample of incoming opportunities might evaluate every one. An account manager might walk into every client call with the relevant history, not just the calls they had time to prepare for.

Those projects may raise total spending. The question is whether the new capability earns its cost. Measure what would make it valuable, such as more qualified opportunities found or fewer missed commitments. Record what happens today, then test the improvement.

When to stop

Before results arrive, agree on three decisions and a review date. For the invoice example, the day-60 review might read:

Expand if accepted invoices cost no more than \$4 each, accuracy matches or beats today's, and the team has used the time to shrink the dispute backlog.

Keep testing if quality meets the standard but cost sits between \$4 and \$6, and only if the team can name a plausible fix, with its own budget and deadline.

Stop or redesign if cost stays above \$6, accuracy falls below today's, or checking the output eats the time the project was supposed to return.

Setting the lines in advance matters because smart people fool themselves in predictable ways once results arrive. They treat the first result as the answer. They say “we’ve already spent too much to stop now.” They believe good news at once and explain bad news away.

New evidence can justify moving a line. Record why it moved and who approved it. Otherwise the target moves forever.

If nothing could make you stop, it isn't a bet. It's a commitment you haven't admitted to.

What expires

Your assumptions can go stale even when you leave the workflow alone.

The product changes. A new version, a different default setting, or a revised feature can change cost and quality. One leading AI vendor's newest version quietly runs at a lower default setting than the last. Check representative work before relying on old measurements.

The market changes. Prices move, and cheaper tools become capable of doing part of the job. A design that was economical six months ago may deserve another look.

The terms change. Salesforce has said it will start charging when registered AI agents connect to its platform. The rate is still listed as “to be announced,” with thirty days' notice promised before billing begins.

Your own work changes. Volumes grow. Customers expect something different. A process that saved time at one scale becomes a bottleneck at another.

Set a regular review date, and name the changes that should trigger an earlier one. Then pull out the page from section 4, compare it with what you now know, and decide whether to keep, expand, redesign, or stop.

The decision was right when you made it. That doesn't make it right now.

Where you stand

Name one person accountable for keeping all of this current. A person, not a committee.

They don't need to own every tool or run every measurement. They need access to spending and workflow data, cooperation from the people doing the work, and the standing to bring renewal, expansion, redesign, and stop decisions to whoever can make them. Hand someone the spreadsheet without those and you've created a reporting task, not accountability.

At a 200-person company, nobody's job is to catch this by default. Someone has to be named, and they should be able to answer four questions: What are we spending? What does an acceptable result cost? What is it worth? What changed because we learned that?

Here are the rungs. Pick the highest one you can support with evidence.

-
- 1 We can't reliably say what we spend on AI.**
 - 2 We know what we spend, but not what an acceptable result costs.**
 - 3 We know the cost, but not what the work is worth.**
 - 4 We know both, and it has never changed a decision.**
 - 5 We write down what we expect, act on results, and revisit when conditions change.**
-

You may be on different rungs for different workflows. Start with one where the spending or the business consequence is material. Give it an owner, a unit, a baseline, and a review date.

ABOUT THIS BRIEF

I build AI systems for organizations end to end, and I'm a Claude Certified Architect. Much of the work is tracing how a result actually gets produced: the tools, the connections between them, the checking, and the decisions underneath.

The aim is to give a company a sound reason to invest, a way to see whether the investment is working, and the ability to change course.

Tell me which rung you're on — reply to the email that brought you this, or write me directly. I'll tell you what I'd look at first.

If a closer look would help, I offer a fixed-scope, one-week audit for \$3,000. It covers the tools you're paying for and how they connect, two workflows traced end to end, and prioritized findings with a scoped first project. We can talk first about whether that would answer the question you actually have.

The brief is yours to keep whether or not we talk.

Scott Jennings

scott@sufficiently-advanced.ai

Sources, verified September 30, 2026. Accenture, "A CIO's guide to AI tokenomics" (September 10, 2026): 750 senior executives at enterprises with more than \$1 billion in annual revenue across 17 countries, plus 15 interviews, June–July 2026. The spending, usage and budget findings concern usage-based ("token") spending, not whole AI budgets; subscriptions and embedded features deserve the same questions. "More than 40 percent" is the author's arithmetic on Accenture's two forecasts ($1.78 \times 0.81 \approx 1.44$), not an Accenture figure. The 90 percent refers to tasks Accenture found did not require the most capable ("frontier") models.

Artificial Analysis, Claude Opus 5.5 evaluation (September 22, 2026): per-token prices cut 20 percent; at maximum effort on its Intelligence Index benchmark, "level with Opus 5 on cost per task despite 1.6x the output tokens." A benchmark result, not a forecast for any company's workflows. Default-setting change: Anthropic, Opus 5.5 migration guide. Agent charges: Salesforce Help, "Understand Alforce Impact" (September 17, 2026), rate listed as TBA and metering not yet begun. The invoice pilot is illustrative; all its costs, volumes, targets and thresholds are hypothetical.