

CFO PLAYBOOK, 2027

Ten questions finance should ask about AI spend, and **what a defensible answer sounds like.**

You will be asked what the AI returned, not what it cost. The cost is the easy half, and most finance teams can produce it. The return needs a record that splits experimentation from production, attributes every production dollar to the team, application and customer that caused it, and ties to the provider invoices you actually paid. This playbook is the build plan for that record, and the ten questions that show whether you have it.

WRITTEN FOR

The CFO, VP Finance or head of FP&A at a company spending real money on AI. Internally on staff tools, copilots and APIs, and for software companies, inside the product as well. You are the person the board turns to when it asks whether the spend is working.

ALSO FOR

The controller who closes the books on it, the audit committee chair who signs off on how it is reported, and the CEO who has to defend it to investors. Each can read the questions to know what to ask finance.

NOT FOR

The engineering leader who instruments it. That is the companion governance playbook, and for software companies the AI-native CTO playbook. This document describes what published sources say. It is not accounting or investment advice for any particular company.

The board stopped asking what AI costs. It is asking what AI returned, and it is asking you.

Two years ago AI was a line inside the engineering budget. It is now large enough to move gross margin, volatile enough to break a forecast, and visible enough that further funding depends on an answer finance has to give. Almost every finance team can produce a total. Very few can produce a return.

66%

of boards now condition further AI funding on proof of return. 43% of finance leaders are already being asked for a return number they cannot produce.

CloudZero, June 2026

22%

of finance leaders can tie AI spend to business results today. 87% say they need that ability within the next year.

CloudZero, June 2026

11%

of organizations forecast AI spending within 10%, down from 15% the year before. 81% cannot fully account for their AI costs.

Mavvrik and Benchmarkit, July 2026

43%

of companies track AI cost by customer. Fewer than a quarter track it by transaction.

CloudZero with Benchmarkit, February 2026

Read the third card twice. Forecast accuracy got worse in a year when spend and attention both grew. That is not a competence problem. It is what happens when two kinds of spend with opposite behaviour are forecast as one number. The next page separates them, and most of this playbook follows from that one move.

There are two pools of AI spend, and only one of them should be predictable.

POOL ONE

Spend that should be volatile

Research, experimentation, prototypes, a team trying a tool, an engineer testing a model against a sample. The variance here is healthy. It is how the company finds what works, and forecasting it tightly is how a finance team teaches engineers to hide spend.

Budget it as an envelope, forecast it loosely, and know who owns it. That is enough.

POOL TWO

Spend that must be predictable

Anything the company has priced, forecast, or promised to a customer. AI inside the product belongs in cost of goods sold. AI that staff depend on to deliver the actual work belongs in a committed operating line. Either way it needs to be predictable, attributed to the team, application and customer that caused it, and checked against the invoice every month.

This is the pool the ten questions are about.

Most finance teams forecast both pools as one number, and that single number is why so few can forecast within 10%. A volatile pool and a predictable pool averaged together produce a forecast that is wrong in both directions. It also hides the moment a research project quietly becomes a production feature that customers depend on and the company now pays for at scale.

The same split answers the board. Pool one is a learning budget, and its return is measured in decisions. Pool two is a cost of delivering something, and its return is measured against what that something earns. The useful question is never what is our AI spend. It is which pool did this dollar come from, what did it return, and can we prove it.

TOKENS ARE THE DENOMINATOR, NEVER THE STORY

Report the ratio of attributed cost to attributed return. It survives every price change.

Token volume and price per token are the inputs engineering watches. They are the wrong unit for finance, because both move without anything changing in the business. On 30 July 2026 one major provider cut a model's price by 80%. A month later another provider's introductory pricing ended and its rates went back up by 50%. A finance team that reports cost per token reports noise.

What changes under you

The provider, the model, the price per million tokens, and how many tokens the same task consumes. An ICONIQ survey of AI builders found 68% found the true cost of internal AI difficult to predict, and one workflow projected at \$0.10 a run reached \$1.50 or more as agents retried.

CloudZero and Digital Applied pricing trackers, August 2026. ICONIQ Growth, July 2026

What holds still

Cost per customer against revenue per customer. Cost per completed task against what the task is worth. Cost per team against what the team ships. A ratio of attributed cost to attributed return carries through a model swap, a price cut and a price rise, which is why it is the number a board can compare quarter to quarter.

THE FINANCE VIEW OF AN AI LINE ITEM

Six moves that turn an AI total into a number you can defend.

None of these needs a new system on day one. Each needs an owner, a cadence and a written rule. Engineering produces the attributed record. Finance decides how it is classified, checks it against the invoices, and reports it.

01 Separate the two pools in the chart of accounts

Open distinct accounts, or a distinct dimension, for experimentation spend and production spend. Product AI lands in cost of goods sold. Committed staff tooling lands in its own operating line. Write the rule that moves a workload from pool one to pool two, usually the day it is priced, forecast or promised to a customer, and name who applies it.

02 Attribute cost to team, application and customer

Allocation by headcount or licence count is a guess wearing a spreadsheet. Ask engineering for cost attributed at the level of each model call, rolled up to the team, the application and, for product AI, the customer. Anything that cannot be attributed goes on its own line named unattributed, so the gap is visible instead of spread.

03 Tie the attributed record to each provider invoice, monthly

A computed figure has no way of knowing it is wrong. Each month, finance compares the attributed total to every provider invoice, line by line, and names every gap as one of two things: usage nobody was tracking, or a price that changed. A gap with a name is a finding. A gap without one is an error in the board pack.

04 Set the forecast method per pool

Pool one gets an envelope with a ceiling and a monthly look. Pool two gets a driver-based forecast: volume of the unit customers buy, times attributed cost per unit, with a sensitivity for provider price. Report forecast against actual for each pool separately, so a variance can be traced to the pool and the driver that caused it.

05 Define the return metric per workload

Every production workload gets one return metric agreed with its business owner before the spend scales. Revenue per customer for a product feature. Cost of the human work it replaced or accelerated for an internal tool. Write the metric down with a baseline and a date. A return defined after the spend is a rationalisation.

06 Report monthly to the board

One page. Spend by pool, against forecast. The tie-out to invoices with gaps named. Cost to return per production workload, with trend. Customers or workloads below their margin floor. The owner of each. Sent on a cadence, not assembled when someone asks.

WHAT THE BOARD WILL ASK, AND WHAT YOU NEED IN HAND

Every board question about AI maps to a piece of evidence finance should already hold.

THE QUESTION YOU WILL HEAR	THE EVIDENCE THAT ANSWERS IT	WHERE IT COMES FROM
What did AI return?	Cost to return for each production workload, against the metric agreed before the spend, with six months of trend.	Moves two and five
Is AI eroding our gross margin?	The production pool inside cost of goods sold, cost per customer against revenue per customer, and the customers below the margin floor.	Moves one and two
Can you forecast it?	Forecast against actual for each pool separately, for the last three quarters.	Move four
Why were we over last quarter?	A variance bridge: volume, provider price, untracked usage, and workloads that moved from pool one to pool two.	Moves three and four
Are these numbers right?	The monthly tie-out of the attributed record to each provider invoice, with every gap named and the largest one explained.	Move three
What if a provider raises prices?	Share of production spend by provider, and the gross margin effect of a material increase from the largest one.	Moves two and four
Who is accountable?	A named owner per pool and per production workload, and the standing monthly report.	Move six

Ask in this order. The first two decide whether any of the others can be answered.

Each question comes with what a defensible answer sounds like, what a plausible answer sounds like, and the follow-up that separates them. The plausible answers are not dishonest. They are what a competent person says when the instrumentation cannot do better, and they are the answers that turn into a board action item with finance's name on it.

QUESTION 01 ASK FP&A

What did AI cost us last month, in dollars, split into experimentation and production?

The baseline, and it is wider than the API bill. AI-native application spend grew 108% in a year, ChatGPT is now the most expensed application, and 78% of IT leaders were hit with unexpected charges from consumption-based or AI pricing. GitHub moved every Copilot plan to usage-based billing on 1 June 2026. A large share of AI cost now arrives through subscriptions, cards and expense reports rather than a provider invoice.

A DEFENSIBLE ANSWER

Two figures, one per pool, covering provider APIs, subscriptions with AI switched on, and expensed tools, with the rule used to place each one.

A PLAUSIBLE ANSWER

One figure taken from the provider invoices. It is accurate as far as it goes, and it misses the subscriptions and the cards.

THEN ASK

Which workloads moved from experimentation to production this quarter, and who decided?

QUESTION 02 ASK THE CONTROLLER

Does our attributed cost record tie to each provider invoice, and what were last month's gaps?

The most important question in the playbook. Usage logs produce a computed figure. The invoice is what was paid. Only 13% of organizations report even basic visibility into their AI spend, and 81% cannot fully account for their AI costs. The gap between the computed record and the bill is where untracked usage and price changes hide, and it is the first thing an auditor will test.

A DEFENSIBLE ANSWER

A monthly tie-out performed by finance, line by line per provider, with every gap named as untracked usage or a price change, and the largest gap in the last six months explained.

A PLAUSIBLE ANSWER

"The dashboard tracks it in real time." A dashboard is a calculation. It is not a tie-out, and it cannot tell you when it is wrong.

THEN ASK

When did the record last fail to match an invoice, and what was the cause?

QUESTION 03 ASK ENGINEERING

Can we attribute every production dollar to a team, an application and a customer?

The attribution question, and the one that turns a total into an instrument. 43% of companies track AI cost by customer, and fewer than a quarter by transaction. Organizations estimate 26% of AI spend is wasted, and waste cannot be found in a total. It is found in the team, application or customer that produced it.

A DEFENSIBLE ANSWER

Cost attributed at the level of each model call, rolled up by team, application and customer, with the unattributed remainder shown as its own line and shrinking.

A PLAUSIBLE ANSWER

An allocation by headcount, seats or revenue share. It sums correctly and tells you nothing about which customer or team drove the cost.

THEN ASK

What share of last month's production spend is unattributed, and what is the plan to close it?

QUESTION 04 ASK FP&A

How do we forecast each pool, and how close were we last quarter?

Only 11% of organizations forecast AI spending within 10%, down from 15% a year earlier. 79% of enterprises overspent on AI in the past twelve months. Forecast error in one blended number cannot be diagnosed, because nobody can say whether it came from healthy experimentation or from a production workload that scaled faster than plan.

A DEFENSIBLE ANSWER

An envelope for pool one, a driver-based forecast for pool two built on volume times attributed unit cost, and forecast against actual reported for each pool separately.

A PLAUSIBLE ANSWER

Last quarter's total plus a growth rate. It is a reasonable starting point and it will be wrong in a way nobody can explain.

THEN ASK

Of last quarter's variance, how much came from volume, how much from price, and how much from a workload changing pools?

QUESTION 05 ASK PRODUCT AND ENGINEERING

What does each production workload cost per customer, or per completed task, and which ones are below the margin floor?

AI cost concentrates, so an average cost per customer hides the tail you are asking about. ICONIQ's survey of companies building AI products puts AI product gross margin at 45% in 2025, projected at 53% for 2026, well below the margins investors are used to from software. AI has eroded gross margins at four in five enterprises for a second year.

A DEFENSIBLE ANSWER

Cost per customer against revenue per customer, or cost per completed task against its price, per workload, with the named accounts below the floor and the spread between cheapest and most expensive.

A PLAUSIBLE ANSWER

"Margins are healthy overall." That is an aggregate statement about a per-customer problem.

THEN ASK

Is any customer unprofitable to serve today, and would we know next month if one became so?

QUESTION 06 ASK THE CEO AND PRICING

Are we charging a flat price for anything whose cost underneath is usage-based?

The classic trap. A flat price set before anyone knew what heavy use looked like turns the heaviest users into the least profitable ones. 49% of organizations were forced to reprice AI-powered products because of unforeseen cost increases. Repricing after launch is slower and more expensive than modelling the break-even point before it.

A DEFENSIBLE ANSWER

A list of products with that shape, the usage level at which each one breaks even, and the number of customers already past it.

A PLAUSIBLE ANSWER

"Most customers are light users." Probably true, and it says nothing about the customers who are not.

THEN ASK

Is the number of customers past break-even growing, and what is the pricing response if it is?

QUESTION 07 ASK ENGINEERING

If usage doubled tomorrow, would finance know before the invoice?

Surprise is the failure mode finance owns. 72% of organizations had an unexpected AI cost spike or bill in the past year, a third more than once, and only 20% could identify the reason within hours if spend doubled overnight. 40% of organizations had an unexpected AI cost that required board-level escalation. A monthly review finds the spike after it has been paid for.

A DEFENSIBLE ANSWER

A named alert on production spend, a threshold per workload, and a named person in finance who receives it the same day.

A PLAUSIBLE ANSWER

"We review spend monthly." That is a reconciliation schedule, not a tripwire.

THEN ASK

What was our largest surprise in the last twelve months, and how did we find out?

QUESTION 08 ASK FINANCE AND PROCUREMENT

What happens to gross margin if our largest provider changes its price?

Concentration risk, and it runs both ways. Within 32 days in mid 2026 one provider cut a model's price by 80% and another's rates went back up by 50% when introductory pricing ended. 61% of IT leaders cut projects because of unplanned cost increases. A cut is only a saving if someone notices it and the forecast reflects it.

A DEFENSIBLE ANSWER

Share of production spend by provider and by model, and the modelled gross margin effect of a material increase from the largest one.

A PLAUSIBLE ANSWER

"We could switch providers." Possibly true, and silent on how long it would take and what it would cost to re-test.

THEN ASK

When a provider last cut a price, how long did it take to show up in our forecast?

QUESTION 09 ASK EACH BUSINESS OWNER

For each production workload, what is the return metric, and what did it return last quarter?

The question the board is actually asking. 78% of finance executives cannot fully tie AI spend to business outcomes. Only 15% of finance leaders can calculate AI return without significant obstacles, and attribution gaps and disagreement between finance and engineering over definitions are among the top reasons. Only 26% of organizations report a sound method for measuring the business value of AI spend.

A DEFENSIBLE ANSWER

One metric per workload, agreed with its owner before the spend scaled, a baseline and a date, and the ratio of attributed cost to attributed return with its trend.

A PLAUSIBLE ANSWER

Adoption figures, or a productivity estimate from a survey of users. Both are signals. Neither is a return.

THEN ASK

Which workload would we stop funding today if we had to cut one, and what number tells us that?

Who owns this number, and what goes to the board every month?

Close on accountability. 52% of organizations say there is no clear owner of AI cost, with responsibility split across engineering, FinOps, finance and IT. DoIT found AI spend responsibility claimed by technology at 55% of companies and by finance at 53%. A number with two partial owners is a number nobody maintains.

A DEFENSIBLE ANSWER

A named owner in finance for the reported number, a named owner in engineering for the attributed record, and a one-page monthly report to the board on a fixed date.

A PLAUSIBLE ANSWER

"Finance and engineering own it together." Shared ownership of a number is unowned.

THEN ASK

If either owner left next month, would the report still go out on time?

READING THE ANSWERS

Count the defensible ones, not the confident ones.

Score an answer as defensible only if it came with a document, a name, or a date attached. Confidence is not evidence, and in a board pack it is a liability.

EIGHT TO TEN

Ahead of nearly everyone.

The remaining gap is usually question nine, because a return metric needs a business owner to commit to it before the spend scales. Close it and the board conversation moves from cost to allocation.

Put the one-page report on a fixed date and send it before anyone asks.

FOUR TO SEVEN

Normal, and worth a decision.

Do not try to close every gap at once. Questions one, two and three are the foundation. Without the pool split, the tie-out and attribution, the other seven are estimates.

Put a name and a date against those three, and start the monthly tie-out this close.

THREE OR FEWER

An instrumentation problem, not a competence one.

Most of the market sits here. The risk is not a careless team. It is a board that discovers the gap first, on its own timetable, in the quarter funding is decided.

Start with the pool split. It changes the answer to four of the ten within a month.

One caution. Measurement surfaces overruns before it prevents them. DoIT found mean overspend of 30.9% at mature FinOps organizations against 16.1% at early-stage ones, because the mature ones could see it. Tell the board before you show the first quarter.

TAKE THIS INTO THE ROOM

AI spend. Ten questions for finance.

Print this, or send it to Engineering, Product and the CEO before the meeting so everyone is asking from the same sheet.

- | | | |
|----|--|----------------------|
| 01 | What did AI cost us last month, in dollars, split into experimentation and production? | FP&A |
| 02 | Does our attributed cost record tie to each provider invoice, and what were last month's gaps? | CONTROLLER |
| 03 | Can we attribute every production dollar to a team, an application and a customer? | ENGINEERING |
| 04 | How do we forecast each pool, and how close were we last quarter? | FP&A |
| 05 | What does each production workload cost per customer, or per completed task, and which are below the margin floor? | PRODUCT, ENGINEERING |
| 06 | Are we charging a flat price for anything whose cost underneath is usage-based? | CEO, PRICING |
| 07 | If usage doubled tomorrow, would finance know before the invoice? | ENGINEERING |
| 08 | What happens to gross margin if our largest provider changes its price? | FINANCE, PROCUREMENT |
| 09 | For each production workload, what is the return metric, and what did it return last quarter? | BUSINESS OWNERS |
| 10 | Who owns this number, and what goes to the board every month? | CEO |

Score an answer as defensible only if it arrived with a document, a name, or a date attached.

WHERE THIS CAME FROM

We build the attributed record that finance works from.

Crystil is an independent third party, SOC 2 compliant. One line of code prices every model call by input, cached and output tokens and the model used, and attributes the cost to a customer, organization or team, with an optional department beneath it. Workflows group many model calls into one task, so a whole outcome has one cost. Customers compares cost per customer with revenue per customer to give margin per customer. Pulse routes a team's AI coding tools, Claude Code, Codex and Gemini CLI, through Crystil with no SDK.

What we do not do matters as much. Crystil does not reconcile your invoices. We provide the attributed cost record your finance team reconciles against its provider invoices. We do not choose your return metric or certify your numbers. Questions three, five and seven are the ones we exist to help answer, and the record behind question two is the one we keep running.

[Talk to an AI measurement expert](#)

crystil.com

COMPANION GUIDES

Forward this to the people who produce the record. The governance playbook is written for the technology leader who owns internal AI. For software companies, the AI-native CTO playbook covers attribution by customer and workflow inside the product.

Sources

1. CloudZero, Finding the ROI of AI: The Finance Perspective, press release, 24 June 2026. 260 senior finance professionals, most of them CFOs. Vendor research.
2. Mavvrik with Benchmarkit, 2026 State of AI Cost Governance Report, press release, 29 July 2026. 396 enterprises, surveyed April and May 2026. Vendor research.
3. CloudZero with Benchmarkit, FinOps in the AI Era: A Critical Recalibration, press release, 12 February 2026. Vendor research.
4. Harness, 2026 State of AI in FinOps, press release, 29 July 2026. 700 engineering leaders in five countries, fielded by Sapio Research. Vendor research.
5. DoiT, Why 79% of Enterprises Overspent on AI in 2026, 9 June 2026. 500 US and UK finance leaders, fielded February 2026 by Sapio Research. Vendor research.
6. Zylo, 2026 SaaS Management Index, 29 January 2026. Platform data and IT leader survey. Vendor research.
7. GitHub, GitHub Copilot is moving to usage-based billing, 27 April 2026.
8. ICONIQ Growth, 2026 State of AI: The Builder's Economy, July 2026. About 300 executives at software companies building AI products.
9. Digital Applied, AI API Pricing, August 2026: Cuts, Promos, and Traps, 5 August 2026, with the CloudZero pricing tracker.

Crystil. v1.0, 7 October 2026. Every figure is published and dated. Nothing here is modelled, estimated or invented. This document describes what published sources say. It is not legal, tax, accounting or investment advice.