

2026 STATE OF

AI Cost Governance

AI adoption has outrun the financial discipline built to govern it. A study of 396 enterprises.

Why are half our Copilot seats unused?

How did we miss the AI forecast again?

We're losing margin on this product.

Who approved this?

We rolled out code assist org-wide and had no idea what it would cost.

Which team owns this spike?

Can we swap to a cheaper model?

Why did our Claude spend jump 3× this week?

Contents

Introduction	04	Theme 4: Financial Impact	16
Research overview	05	Theme 5: Cloud & AI Infrastructure	18
The three patterns of AI cost	06	Theme 6: Developer Tooling & the SDLC Gap	19
Top findings	07	Theme 7: Industry Insights	20
Theme 1: AI Cost Governance	09	What's Next: Role-Based Opportunities	22
Theme 2: Agentic AI	11	Conclusion	27
Theme 3: Cost Visibility & Attribution	14	About	28

The bill has arrived

Enterprise AI spent two years in experimentation: pilots, proofs of concept, innovation budgets with loose accountability. That phase is over.

AI costs now influence budget reviews, pricing decisions, product investments, and board discussions. Across **396 organizations** surveyed in April and May 2026, AI adoption has pulled away from AI cost visibility. AI programs are running on a financial picture that is, at best, incomplete and, at worst, misleading.

Coverage, not effort

More organizations track AI costs, yet forecast accuracy **declined**. Tracking is only as useful as what it covers, and the fastest-growing categories, agents, GPU infrastructure, on-prem, sit largely outside the frameworks doing the reporting. You cannot predict what you cannot see.

Bill shock is mainstream

62% report that unexpected AI costs materially altered a business decision this past year: emergency freezes, forced repricing, board-level escalation. **One in four** delayed or cancelled an AI initiative outright.

ROI is premature

ROI is math: $(\text{Value} - \text{Cost}) / \text{Cost}$. It needs a denominator. With an incomplete cost picture, that number is missing. Before anyone can answer whether AI delivers a return, they need confidence in what it costs.

THE GAP: THEY TRACK, THEY CAN'T FORECAST

98%

Track AI infrastructure costs

95%

Assign formal AI budgets

11%

Forecast within $\pm 10\%$ accuracy

98% track. 11% forecast. The gap between those two numbers is the report.

Research overview

This research, conducted by Mavvrik with **Benchmarkit** and Ray Rike, surveyed 396 organizations across six sectors during April and May 2026: Technology/SaaS (including B2B SaaS, Vertical AI, AI-Native App, and AI Infrastructure, 47% combined), Financial Services (23%), Manufacturing (9%), and Retail (6%).

The 2026 edition goes deeper than 2025. Last year's research documented the problem; this year's measures how much further AI spend has pulled ahead of the discipline meant to govern it, and the distance widened on every front: forecast accuracy declined year over year (15% to 11%), hybrid adoption grew (61% to 68%), and agentic workloads reached near-universal deployment while attribution stayed stuck at the same department level it held in 2025.

The most important finding is one of breadth and depth. Almost every organization now tracks AI costs; the challenge is how far that coverage extends, with visibility falling off sharply as spend expands into agents, GPUs, developer tooling, hosted models, data platforms, and private infrastructure.

396

Organizations
surveyed

6

Industry
sectors

7

Thematic
areas

SEVEN THEMATIC AREAS

01 AI Cost Governance

03 Cost Visibility & Attribution

05 Cloud & AI Infrastructure

07 Industry Insights

02 Agentic AI

04 Financial Impact

06 Developer Tooling & the SDLC Gap

Three patterns of AI cost

What drives AI cost depends on how it’s deployed. In a customer-facing product, customers drive consumption. In internal workflows, employees do. In agentic systems, agents do. Each model accumulates cost differently, breaks differently, and requires different governance.

	AI in Product	AI in Workflow	Agentic AI
	Customers drive cost	Employees drive cost	Task complexity drives cost
PRIMARY IMPACT	Gross margin, COGS, cost-to-serve	Productivity assumptions, headcount decisions	Forecast accuracy, runtime volatility
OWNER	CFO, Product	Engineering, FinOps	Shared, often undefined
WHAT BREAKS	Margin erosion, forced repricing	Unvalidated ROI, hidden spend	Retry loops, orchestration overhead
KEY METRIC	Cost per inference, per customer	Cost per workflow, per seat	Cost per agent, per task run
GOVERNANCE LEVER	Attribution at customer and feature level	Attribution at workflow and tool level	Instrumentation before scale
THE PROOF	70% have not required COGS-based AI cost tracking	32% lack a single accountable owner for AI infrastructure costs	98% / 36% run agentic workloads / include them in cost reporting

Three drivers. Three governance problems. Each needs to be managed differently.

THE MISCLASSIFICATION PROBLEM

Same cost. Different problem.

Product AI and Workflow AI look similar on an invoice. One scales with customers and hits gross margin. The other scales with employees and accumulates as unvalidated OPEX. The governance response differs: product AI needs attribution at the customer and feature level; workflow AI needs it at the agent and workflow level.

49%

were forced to reprice an AI product after costs moved, usually mid-cycle, in front of customers.

COGS VS. OPEX: WHY IT MATTERS

If it's in your product, it's COGS.

Every session, output, and model call affects gross margin directly. Your pricing model depends on knowing what each unit of AI output costs before the customer consumes it. When that cost shifts unexpectedly, your margin shifts with it.

If it's in your workflows, it's OPEX.

Cost scales with employee activity, not revenue. The risk is not margin erosion – it is invisible spend accumulating against productivity assumptions no one has tested. Attribution stops at the department level, which tells you who spent the money but not why it cost what it did.

Reporting them together produces a number that's accurate in aggregate and useless for decisions.

Top findings

An executive summary: each is examined in full in the sections that follow.



Forecast accuracy is getting worse, not better

Only 11% forecast AI spend within $\pm 10\%$, down from 15% in 2025: 39% miss by 11 to 25%, 35% by 26 to 50%, and 13% by more than 50%.



Unexpected costs are altering strategy

62% say a cost surprise materially changed a business decision last year. **One in four** stopped an AI program outright.



44% rate themselves Advanced. 89% can't forecast.

44% of organizations rate themselves Advanced in AI cost management -- the same population where 89% cannot forecast within $\pm 10\%$. Maturity ratings reflect inputs (budgets, tools, teams) while the outcomes those inputs exist to produce stay out of reach.



AI costs are eroding gross margins

81% report moderate-to-high gross-margin impact: 63% see a 6 to 15 point decrease, and 18% lose more than 16 points.



The fastest-growing categories have the least visibility

The cost categories growing fastest, agentic workflows, GPU infrastructure, and on-premises environments, have the least governance coverage. Organizations are expanding into exactly the environments where their current tracking frameworks break down.



Agentic AI scaled before attribution

98% run agentic workloads; **15%** can't attribute those costs at any level: a gap 5× wider than for general AI (3%).



Developer tool costs are invisible to finance

98% use AI coding tools, averaging **2.4** per org. 39% exceed expected cost; most finance teams can't see the combined footprint.



Hybrid is the default

68% run hybrid workloads, up from 61%. Third-party hosted LLMs jumped **19% → 32%**, the largest single-category gain in the survey.

TRACKING WENT UP. ACCURACY WENT DOWN.

AI cost tracking is now the baseline, yet forecasting it accurately is rare. Reporting stops where the fastest-growing costs begin: agents, GPUs, on-prem.

— 2026 STATE OF AI COST GOVERNANCE

AI cost governance

Budget discipline has outrun visibility. AI shows up in every budget now, but the structures meant to manage it haven't kept up.

Tracking is not the same as governing

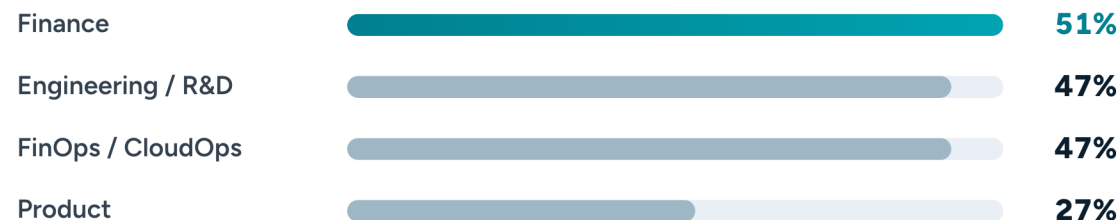
98% track AI infrastructure costs, and **95%** assign a formal budget. That puts AI inside the cost-management mandate at nearly every organization here. What tracking and budgeting can't do on their own is produce accountability: who owns the cost, at what level of detail, with enough clarity to act when something goes wrong.

Visibility is the primary challenge, not process

Lack of cost visibility ranks as the **top** AI infrastructure challenge overall, ahead of inaccurate forecasts and consumption-based pricing. All three trace back to the same thing: the signals exist across every environment, but nothing collects them in one place.

Cloud-era visibility was a property of the platform. AI cost visibility has to be built.

Three functions claim the same dollar



Sums past 100%: ownership is shared or contested across functions. 17% say it varies by use case, 14% describe shared responsibility with no clear owner, 1% have none at all.

In practice, when costs spike in a fragmented ownership environment, no single function is positioned to act before the bill arrives, which is why 40% of organizations in this survey required board-level escalation.

32%

lack a single accountable owner for AI infrastructure costs. When costs spike in fragmented environments, no one is positioned to act before the bill arrives.

THE MATURITY PARADOX

Advanced, by what measure?

Self-assessed maturity scores measure what organizations can see. The problem is what they cannot. Organizations that rate themselves Advanced are rating their inputs: budgets assigned, tools installed, teams named. The output that matters is forecast accuracy, and 89% of organizations miss it.

- ✓ **Advanced** ratings reflect inputs (budgets, tools, teams)
- ✓ Forecast accuracy is the **outcome** those inputs exist to produce
- ✓ The blind spots that should lower the score are the same ones that hide from the scorer

MATURITY SELF-RATING VS. FORECAST ACCURACY


44%

Rate themselves **Advanced** in AI cost management

11%

Actually forecast AI spend **within ±10%**

The same population. Advanced ratings reflect budgets assigned, tools installed, and teams named, while the outcome stays out of reach for nearly nine in ten.

Agentic AI

98% are running agents. 36% report what those agents cost. Adoption got to near-universal before anyone built the instrumentation to govern the spend.

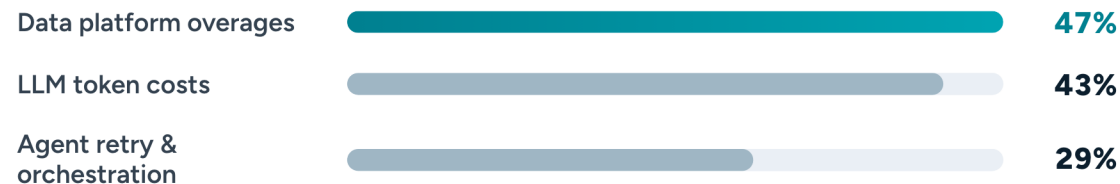
Cost patterns unlike anything before them

One agent task can fan out into a dozen model calls, data-platform queries, and retry loops, with some steps failing quietly and retrying at extra cost. Run the same workflow twice and the second run can cost **2x or 10x** more, depending on what the agent runs into. Unlike traditional software, where the same input produces the same compute cost, AI systems are non-deterministic. The same prompt, agent, or workflow can take different paths, make different tool calls, and consume materially different resources at every execution. No per-request cost model sees that coming.

ROI measurement hasn't caught up

Time or labor savings (**57%**) is the primary lens for agentic ROI; revenue impact (**50%**) ranks second as organizations begin to connect agent performance to outcomes. **9%** running internal agents have no ROI measurement at all.

Top sources of unexpected AI spend



The lead insight surprises most readers: data-platform costs beat tokens. Snowflake, Databricks, and BigQuery are now central AI-infrastructure costs.

37%

Only **37% attribute agentic cost down to the agent or model**, the level where tuning happens. Reporting tends to stop at the workflow (56%) or the department (50%).

Agentic AI hit scale before attribution did

15% of organizations running agentic workloads can't attribute those costs at any level. The tracking simply hasn't caught up to a category that barely existed at scale a year ago. Adoption came first, and visibility is still catching up to it.

AGENTIC ADOPTION VS. COST REPORTING



A 62-point gap between running agents and accounting for what they cost.

BUILDING BEYOND THE BILLING API

Cloud-era visibility was a
property of the platform. AI
cost visibility has to be built.

— 2026 STATE OF AI COST GOVERNANCE

Cost visibility & attribution

Coverage ends where the billing API does. The environments with the least visibility are the ones growing fastest.

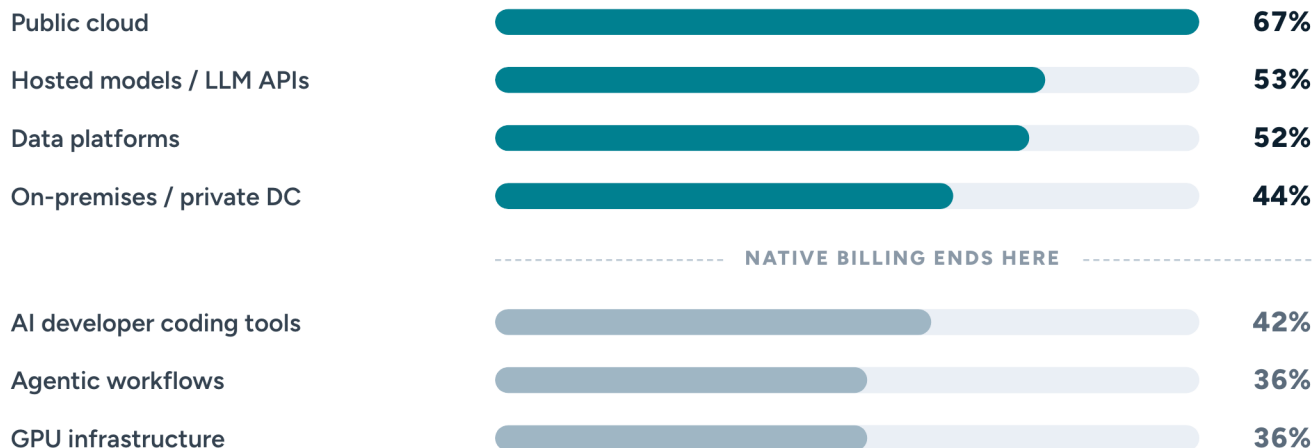
Reporting ends where the billing API does

Coverage is strongest where an environment produces structured billing data, and falls in direct proportion to how far it sits from a native billing API. On-prem GPU clusters, which have **no billing API at all**, climbed from 35% to 44%. Agentic workflows and GPU infrastructure are still stuck at 36%, even though they're growing faster than anything else in the survey. For regulated industries and organizations with data residency requirements, private AI (self-hosted models, on-premises GPU clusters, and local inference) is an additional driver of hybrid adoption that sits almost entirely outside current governance frameworks.

Hybrid has outpaced the tools built to govern it

68% run hybrid workloads, up from 61%. Running AWS (73%), Azure (65%), and GCP (58%) at once is routine, and every one of them has its own pricing model and its own bill. The frameworks doing the governing were built for one thing: reconciling cloud invoices.

AI cost reporting coverage by environment



The drop from 68% to 36% is the story: visibility is strongest where cost comes from a single billing source, and weakest where it is spread across systems.

Consumption pricing has no clean way to forecast it

60% call consumption-based AI forecasting a major or moderate challenge, and only 3% don't use consumption pricing at all. Cloud spend follows fairly predictable infrastructure patterns. AI spend rides on runtime behavior: token consumption, workflow design, retries, model choice. Old-school budgeting can't keep up with that kind of variability.

THE ATTRIBUTION GAP

Attribution follows the org chart, not the cost structure

Team and department attribution come first, and both answer the same question: who spent the money.

Attribution by agent, model, customer, or user, the kind that tells you why it cost what it did, shows up far less. Reporting usually nails the who. The why, the part that lets you cut cost, is where it falls short.

Attribution at the agent, session, and model level requires instrumentation decisions made at deployment time. Most organizations running agentic workloads did not build that tracking in before deployment, and retrofitting it is significantly harder.

AI COST ATTRIBUTION METHODS IN USE

WHO SPENT IT

Team / Department		54%
Product / Feature		53%

WHY IT COST THAT

Agent / Workflow		46%
Model / Provider		43%
Customer		40%
Developer tool user		29%

Knowing which department spent the money is not the same as knowing why it cost what it did.

Unit economics

67% track AI cost-to-serve and use it to make pricing and margin calls. Another 23% track it and never price from it. Producing the number is the measurement phase. Pricing from it is the governance phase. That 23% gap is where AI cost surprises become margin events.

Financial impact

AI bill shock has gone from rare to routine. It's reshaping margins, pulling leaders into board meetings, and killing programs outright.

Forecast accuracy declined year over year

11% forecast AI spend within $\pm 10\%$, down from 15% in 2025. The deterioration is not about maturity: as programs scale and agentic workloads multiply, cost volatility compounds **faster than forecasting models can adapt**.

Most overruns are discovered after the fact

46% have no real-time alert mechanism for AI cost overruns; 63% rely on manual review during reporting cycles, and **35%** discover overruns only after the invoice arrives. For consumption-based pricing, that is not cost management.

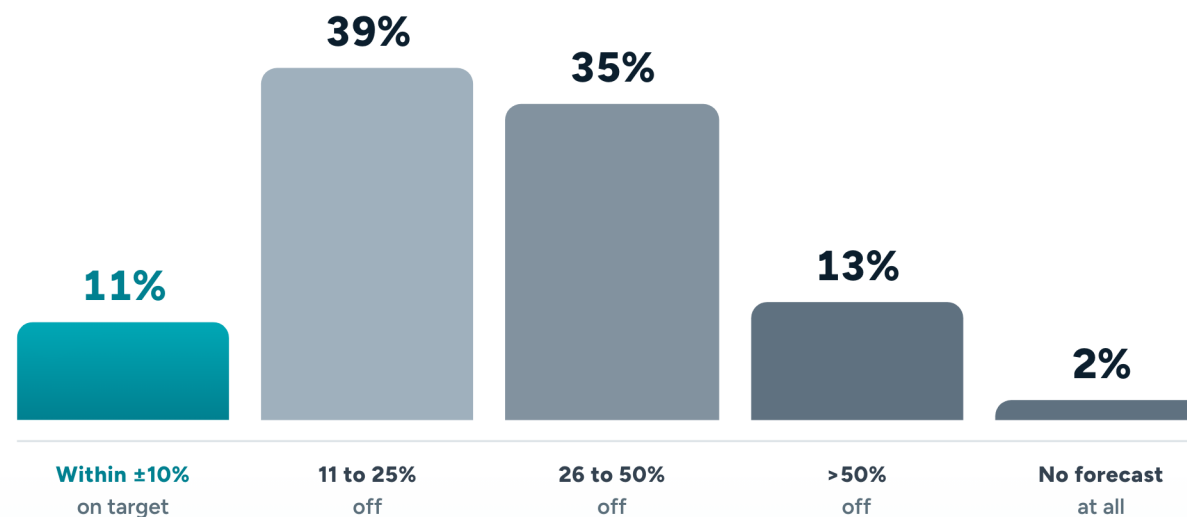
Cost enters the decision too late

60% decide on AI cost before deployment, at the design or architecture stage. 47% only weigh in after deployment, through monitoring. 15% don't see it until the invoice arrives, which means cost never entered the decision in the first place.

Measuring against the wrong denominator

81% report moderate-to-high gross-margin impact. Only **30%** measure AI cost as a percentage of COGS, the metric tied to gross profit, while 56% measure against revenue, which benchmarks spend levels but never answers whether a feature is margin-accretive. The other 70% are making AI investment decisions without the foundational data to evaluate them.

How far organizations miss their AI forecasts



The mass sits in the middle: **74% miss by 11 to 50%**, and 13% miss by more than half. In total, 89% land outside $\pm 10\%$.

70%

have not required COGS-based AI cost tracking: leaving the one number that connects AI spend to gross profit unmeasured.

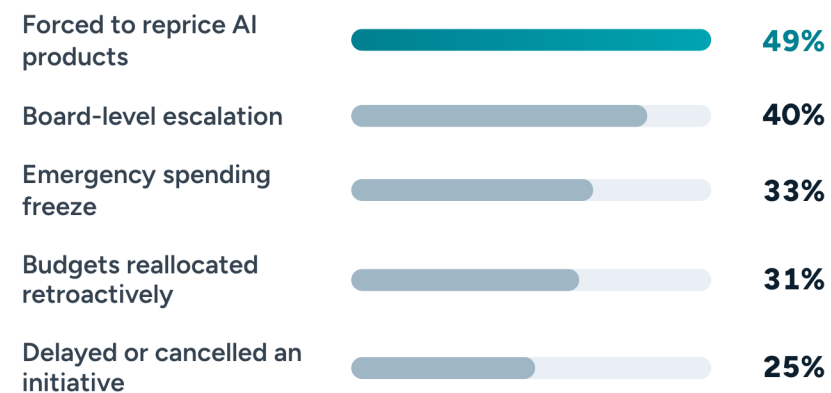
THE COST OF AI BILL SHOCK

One in four stopped the program

62% say an unexpected AI cost materially changed a business decision in the past year: the consequences run from repricing to the boardroom.

- ✓ Repricing mid-cycle **signals instability** to customers
- ✓ Board escalation is **organizational distraction**
- ✓ Cancellation means the investment **produced nothing**

BUSINESS CONSEQUENCES OF UNEXPECTED AI COSTS



These are not FinOps optimization problems. They are enterprise risk-management failures.

Cloud & AI infrastructure

Multi-cloud is the norm. Hybrid is universal. The governance architecture built for cloud billing was not designed for this environment.

Most organizations manage three hyperscaler bills at once

Multi-cloud is rarely a strategic choice. It is the accumulated result of individual workload decisions, vendor relationships, M&A, and geographic requirements, and for finance it means more contracts, more invoices, and cost surfaces that share no reporting standard. Most governance frameworks were built to normalize a single cloud invoice. The organizations in this survey are managing three at once, each with its own contract terms, pricing, and billing cadence. When finance and infrastructure haven't aligned on a unified reporting layer, the result isn't one hybrid strategy, it's three separate ones under a single budget line, with no view of what they cost together.

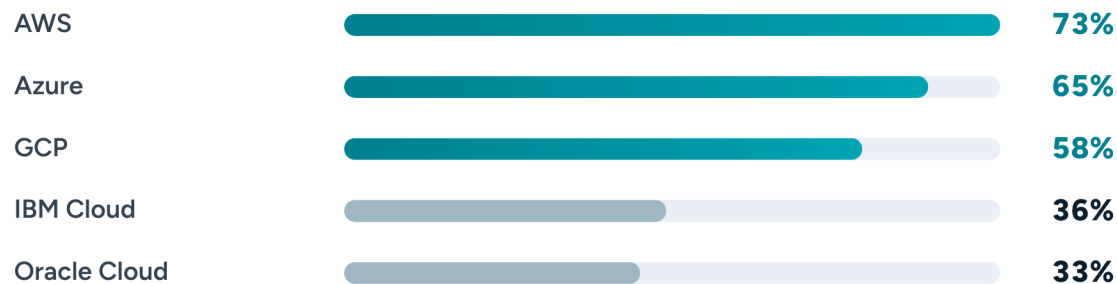
Hybrid is accelerating, not stabilizing

68% run hybrid AI workloads, up from 61%. Only 29% stay entirely in public cloud. Third-party hosted LLMs jumped **19% → 32%**, the largest single-category gain in the survey, and it spans every company size.

The full stack spans build and buy

LLMs via API lead third-party adoption at **83%**, making token consumption the most common variable cost. Data platforms (61%) and GPU providers (48%) complete the stack: each with pricing models that resist unified forecasting.

Cloud provider usage: most run all three



The three-hyperscaler story (73 / 65 / 58) is the lead: most organizations operate all three simultaneously with no shared reporting layer.

47%

name data-platform overages the top unexpected cost, ahead of LLM tokens (43%). A framework that nails tokens but misses the data platform is still a broken forecast.

Developer tooling & the SDLC gap

Developer AI spend is everywhere, yet few are governing it.

AI-assisted development is the new normal

98% of organizations use AI coding tools now. They went from experiment to standard practice a while ago, and the cost accounting never followed. They come in through software licensing, team allocations, and overhead, which keeps them out of view of whoever owns AI cost governance.

Developer costs route around the AI budget

Only **42%** fold AI developer tools into AI cost reporting, and **39%** say the tools cost more than the license or usage they expected. Whatever budget assumption got made at purchase doesn't survive scale.

The footprint grows as agentic coding matures

Today the tools handle autocomplete, generation, and review. Once agents start planning, writing, testing, and shipping with barely any human in the loop, that footprint gets a lot bigger. Instrumenting it now is what makes the consolidation and contract calls possible later.

Are AI developer tools in the budget?



98% use them. 42% account for them.

Nearly every organization runs AI coding tools, yet fewer than half include them in AI cost reporting. **Claude Code (60%), Copilot (54%) and Q Developer (51%)** lead adoption, routed through licensing and overhead, out of finance's view.

■ In cost reporting ■ Outside the budget

2.4

AI coding tools per organization, on average: each with separate billing, separate budgets, and no unified view of the combined footprint.

Industry insights

The same underlying problem looks different depending on what AI is doing, who's accountable, and how solid the governance was before AI showed up. Context changes how that problem surfaces, but it doesn't change whether it's there.

AI-NATIVE APPLICATION

3% forecast $\pm 10\%$

The **worst forecasters in the survey**: less than a third of the already-low 11% overall, despite 52% rating themselves Advanced. More AI means more complexity, faster than governance adapts.

FINANCIAL SERVICES

10% forecast $\pm 10\%$

The most established compliance apparatus of any industry, yet some of the **weakest cost visibility**: 17% can't attribute agentic costs at any level. Compliance discipline never extended to AI cost.

AI INFRASTRUCTURE

39% cancelled

Delayed or cancelled an initiative over cost surprises: **more than 50% higher than the overall rate** and the highest of any segment. The sector powering everyone else's AI has the worst bill shock.

B2B SAAS

25% track % of COGS

AI costs in B2B SaaS should flow directly into COGS, however only 25% measure it that way, least of any segment: 56% benchmark against revenue, which never reveals whether AI features are margin-accretive.

MANUFACTURING

60% forced to reprice

Only 34% saw a decision altered, well below the 62% overall. But where AI meets the customer, **pricing pressure is among the highest in the survey**.

RETAIL · THE BENCHMARK

22% forecast $\pm 10\%$

Double the overall rate, the highest of any segment. Retail concentrates AI in high-volume, repeatable use cases (demand forecasting, inventory, recommendations) where consumption is stable enough to model, and 57% run real-time alerts. Hold new use cases to Retail's measurement bar before scaling, and the accuracy follows.

WHY 11% GET IT RIGHT

The 11% who forecast accurately aren't smarter or better funded. They made a call to push coverage past cloud billing, into agents, GPUs, developer tools, and private infrastructure

01 CFO / Finance Leader

OWNS THE MARGIN

AI spend is now a margin input

The highest-leverage CFO intervention is pre-deployment cost approval. Without it, 15% of organizations in this survey had no cost review before the commitment was made.

These dynamics explain why that control is now a financial governance requirement.

89%

miss their AI forecast by more than $\pm 10\%$.

The gap comes down to coverage.

WHAT TO ACT ON

- ✓ **Require pre-deployment cost approval**
15% of organizations only see AI costs after the invoice arrives, meaning no cost review happened before the commitment was made. Make pre-deployment cost review a standard financial control.
- ✓ **Reporting cycles won't fix the forecast**
Adding more reporting cycles won't move a 9-in-10 miss rate. The forecasting model itself has to change: from carrying fixed costs forward to forecasting by driver, cost per workflow run.
- ✓ **Measure against COGS**
81% report margin impact, but 70% have never required COGS-based tracking. COGS is the one number that shows whether an AI feature is paying for itself. Revenue-based measurement won't tell you that.
- ✓ **AI bill shock is reaching the board**
40% have escalated to the board, 33% froze spend, 25% killed a program. Treat real-time monitoring as a financial control, not a FinOps nicety.

02 CIO / CTO

OWNS THE ARCHITECTURE

Architecture sets tomorrow's visibility

Hybrid is the right architecture for many organizations. It also creates a governance obligation that **most current tooling cannot fulfill**.

36%

include agentic workloads in cost reporting, even as **98% already run them**.

WHAT TO ACT ON

- ✓ **Hybrid needs unified signal collection**
68% run hybrid workloads. On-prem GPU, private LLMs and local data platforms have no billing APIs: only 44% include on-prem in reporting at all.
- ✓ **Agentic deployment outran governance**
Attribution is an instrumentation decision made **at deployment time**, not retrofitted later. Build agent-level cost tracking into the platform before scale.
- ✓ **Lagging detection is not detection**
46% have no real-time alert; 63% rely on manual review. For consumption-based pricing, finding an overage after the invoice isn't management.

03 FinOps / CloudOps Leader

OWNS EFFICIENCY & REPORTING

Signal collection is an operational problem

Cloud billing APIs are necessary but they don't reach everywhere AI runs. The signals are in every environment, and collecting them into one place is your job to own.

46%

have no real-time mechanism to detect AI cost overruns: most still find them in **manual review**.

WHAT TO ACT ON



Attribution must match the decision

Team and department answer **who spent it**. Agent, model and user attribution answer **why it cost what it did**: the layer where governance becomes optimization.



Extend the perimeter past cloud

Agents and GPU infrastructure appear in reporting at just **36%**. Coverage has to reach every environment, billing API or not.



Detection lag is the metric to cut

Move from monthly review toward real-time alerting. **35%** still discover overruns only when the invoice lands: the slowest possible signal.

04 Head of Product / Chief AI Officer

OWNS THE UNIT ECONOMICS

Unit economics decide **what ships**

Product AI and workflow AI are different problems with different owners, metrics, and intervention points. **A program addressing one is solving half.**

49%

were forced to reprice an AI product after costs moved, usually mid-cycle, **in front of customers.**

WHAT TO ACT ON

- ✓ **Price the feature, not the platform**
When cost scales with usage, your most popular feature can become your least profitable without a cost-per-interaction number. 23% of teams produce a cost-to-serve figure and never price from it. Get it to the customer and feature level and you can price with confidence up front, rather than scrambling to reprice mid-cycle.
- ✓ **Build agentic attribution before materiality**
15% of organizations running agents can't attribute those costs at any level. The window to instrument before scale creates exposure is closing fast.
- ✓ **Separate product AI from workflow AI**
The two have different cost drivers and different owners. Roadmap decisions need the unit economics of each, not a blended number that hides both.

05 Engineering Leader

OWNS THE TOOLING

The tooling footprint compounds

AI now writes, reviews, and increasingly operates software. The cost of that shift **routes around the AI budget**, for now.

2.4

AI coding tools per organization on average: each separately billed, **most invisible to finance**.

WHAT TO ACT ON



Move developer tooling into the AI budget

98% adoption with just **42%** finance reporting is an accountability gap. The tools route through licensing and overhead, out of the governance function's view.



Instrument before agentic SDLC scales

As agents plan, test and deploy with less human input, the footprint expands materially. **39%** already exceed expected tool cost.



Treat tool sprawl as a cost surface

Averaging 2.4 tools per org, with separate billing each, consolidation and contract leverage are only possible once the combined footprint is visible.

Visibility first. Everything follows.

AI advancement is happening at record pace, yet as this study shows, adoption is far exceeding proper cost governance. Two years into enterprise AI at scale, fewer than one in nine organizations can forecast what they'll spend before the bill arrives, and the gap is still widening.

Much of this is because the foundation underneath these programs remains incomplete. While the conversation has jumped to AI ROI, the reality is you can't optimize, forecast, attribute, or measure the return on anything without understanding what it costs. The denominator (total cost across every environment) is incomplete for most organizations. The numerator (the value AI actually delivers) is often subjective and difficult to quantify. Both require a commitment, just like companies made with cloud.

From here, it comes down to one question: whether the financial infrastructure gets built ahead of the spend, or only after bill shock makes it unavoidable.

WHAT TO DO BEFORE 2027

These recommendations are sequenced by dependency: visibility is the prerequisite. Forecasting, attribution, and ROI measurement all require complete cost signal coverage as their foundation.



Extend coverage to the full stack

Cloud-only or LLM-only visibility is structurally insufficient. Signal collection must reach **every environment (including those with no billing API)**.



Require forecast accuracy as a governance KPI

Self-assessed maturity predicts nothing. **Forecast variance, attribution coverage, and detection lag** are the numbers that matter.



Treat product AI and workflow AI as separate problems

The metrics, the owners, and the intervention points differ. **A program addressing one is solving half the problem.**



Build agentic attribution before materiality

15% running agents can't attribute costs at any level. **The window to instrument before scale creates exposure is closing.**



Move developer tooling into the AI budget

98% adoption with near-zero finance reporting is an **accountability gap** that widens as agentic SDLC adds another cost layer.



Rebuild the forecasting model

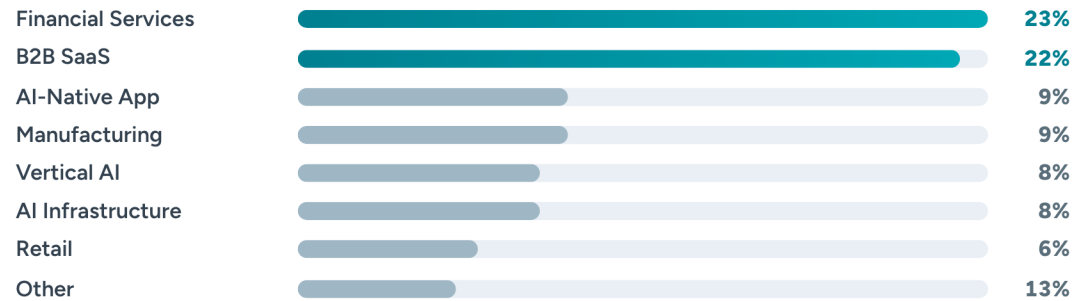
Consumption-based pricing requires **driver-based forecasting**: cost per workflow run, cost per interaction. Fixed-cost extrapolation cannot close the gap.

SURVEY PARTICIPANTS

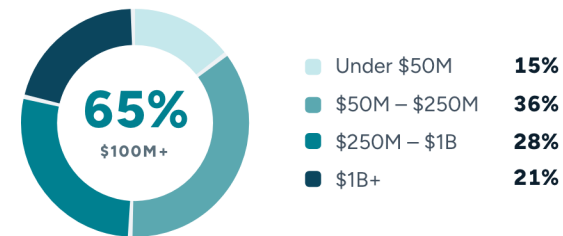
The participants

396 organizations across seven industries, surveyed April to May 2026.

INDUSTRY

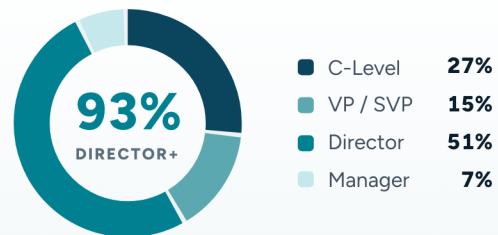


ANNUAL REVENUE



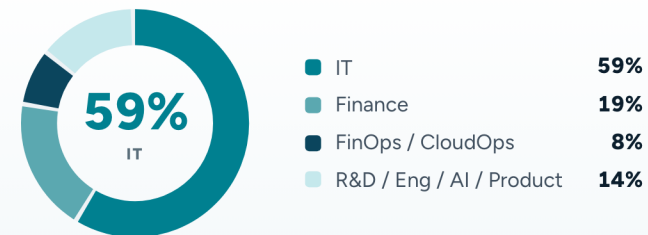
65% of respondents represent organizations with \$100M or more in annual revenue, confirming enterprise-scale participation.

RESPONDENT TITLE



93% are Director level or above, ensuring decision-maker visibility into AI cost governance practices.

DEPARTMENT



Responses span IT, Finance, and FinOps – the three functions this report addresses.

FINANCIAL CONTROL FOR THE MODERN AI ERA

AI didn't create the visibility problem, but it made ignoring it much more expensive.

● ABOUT MAVVRIK

Mavvrik is the financial control center for the modern AI era, providing enterprises with complete **visibility, attribution, and governance** across the full AI stack, including GenAI services, autonomous agents, GPU infrastructure, cloud, SaaS, and private infrastructure. Built for CFOs, FinOps leaders, and IT executives, Mavvrik eliminates AI bill shock and transforms AI costs into strategic investments. For more information, visit mavvrik.ai

● ABOUT BENCHMARKIT

Benchmarkit uses metrics and benchmarks to align B2B SaaS executives across the entire customer journey including customer acquisition, retention, and expansion, leading to better metrics-informed and benchmark-validated decisions. Benchmarkit's vision is to enable every B2B SaaS company to increase revenue growth efficiency and enterprise value by having free access to the most timely and contextual benchmarks available in the industry. For more information visit benchmarkit.ai