RESEARCH

2025 State of AI Cost Governance

mavvrik. benchmarkit

Table of Contents

Introduction

AI Cost Crisis: Finance Leaders Face Margin Erosion and Forecast Chaos

AI infrastructure costs are already reshaping corporate profitability, but most finance teams lack the visibility and control to manage the impact. This research reveals four critical challenges that demand immediate CFO attention:

- **Forecast Failure:** 85% of companies miss AI cost forecasts by more than 10%, with nearly 25% missing by over 50%, creating massive gross margin risk as AI spending scales.
- **Margin Hemorrhaging:** 84% of companies report AI costs eroding gross margins by more than 6%, with over a quarter seeing hits of 16% or more. For example, a product at 80% gross margin could drop to 74% once AI costs are factored in.
- **Visibility Breakdown:** Only 35% include on-premise costs in AI reporting, and half of companies with AI-core products aren't tracking their LLM API expenses—creating dangerous blind spots in cost-to-serve calculations.
- **Infrastructure Complexity:** 61% operate hybrid AI environments spanning public cloud, private infrastructure, and third-party services, fragmenting cost visibility and governance across multiple vendors and billing systems.

The accountability gap is real: Companies charging for AI consistently demonstrate 2-3x better cost discipline than those giving AI features away for free, suggesting that revenue pressure drives the governance rigor most finance teams desperately need.

Research Overview

85%

of companies cannot forecast AI costs within 10%.

For CFOs watching AI expenses balloon while gross margins shrink by double digits, this isn't just a forecasting problem, it's a strategic crisis hiding in plain sight.

The numbers tell a stark story: Across the full sample (N=372), 84% report AI costs eroding product gross margins by more than 6 percentage points (600 bps), with over a quarter seeing hits of 16+ points (1600 bps).

Yet most finance leaders are flying blind: unable to predict next quarter's AI spend, attribute costs to specific products or customers, or even see what's happening across their hybrid infrastructure environments.

This research, conducted by Mavvrik in partnership with Benchmarkit, surveyed 372 companies to understand how organizations are building, running, and financially governing AI workloads. What we found reveals a market in transition: AI has moved from experimental budget line to material cost driver, but the financial discipline hasn't caught up.

The stakes couldn't be higher. As AI transforms from "nice to have" to "must have," the companies that master cost visibility and control will protect their margins while competitors watch profits disappear into untracked infrastructure costs.

***Note: Gross margin impact findings reflect product delivery (COGS). About 70% of respondents were SaaS and AI-native vendors, where inference, GPU, and API costs directly affect gross margin. For enterprises using AI internally, financial impact typically flows through OPEX and operating margin instead.*

Top Findings

AI costs are already eroding gross margins

84% of companies report more than a 6% hit to gross margin from AI costs. Within that, 58% see a 6–15% reduction and 26% report 16%+ erosion. The financial impact is widespread and immediate, making cost visibility and control a strategic imperative for both finance and product leaders.

Forecast accuracy is alarmingly low

Only 15% of companies forecast AI costs within ±10%. A majority (56%) miss by 11–25%, and nearly one in four (24%) miss by more than 50%. For CFOs and budget owners, this level of unpredictability makes it harder to protect gross profit targets as AI grows as a share of COGS.

Hybrid complexity is the default

61% of companies run AI workloads across a combination of public and private environments. This pattern spans all company sizes, including small businesses, and creates greater difficulty in achieving unified cost reporting and governance.

Repatriation is becoming mainstream

67% of companies are actively planning to repatriate some AI workloads to owned infrastructure, and another 19% are evaluating the move. The trend is most active in mid-market companies, while large enterprises are more often in the evaluation stage.

The AI cost surface is broader than tokens

Data platform usage is the #1 source of unexpected AI costs (56%), followed by network access to models (52%). LLM token costs rank fifth (37%). This diversity of cost drivers makes AI spend harder to forecast and control.

Visibility and attribution gaps block action

Only ~35% of companies include on-prem components in AI cost reporting, and about half include LLM API costs even when AI is a core product component. Teams say the #1 tactic to improve cost management is unified visibility across environments; clear cost attribution is #2.

Charging for AI correlates with stronger cost discipline

Organizations that charge or package AI separately are consistently more likely to track cost-to-serve precisely, use real-time usage alerts, and attribute costs by customer, product, or model than those who include AI "for free."

Azure is winning in the enterprise

AWS leads overall cloud usage (77%), but among companies with more than $250M in revenue, Azure adoption climbs to 82%, surpassing AWS in this segment. Google Cloud holds third at 65%, and IBM Cloud maintains niche strength in specific industries.

01 Cloud & AI Infrastructure

Multi-cloud is the new standard, with Azure surging in the enterprise

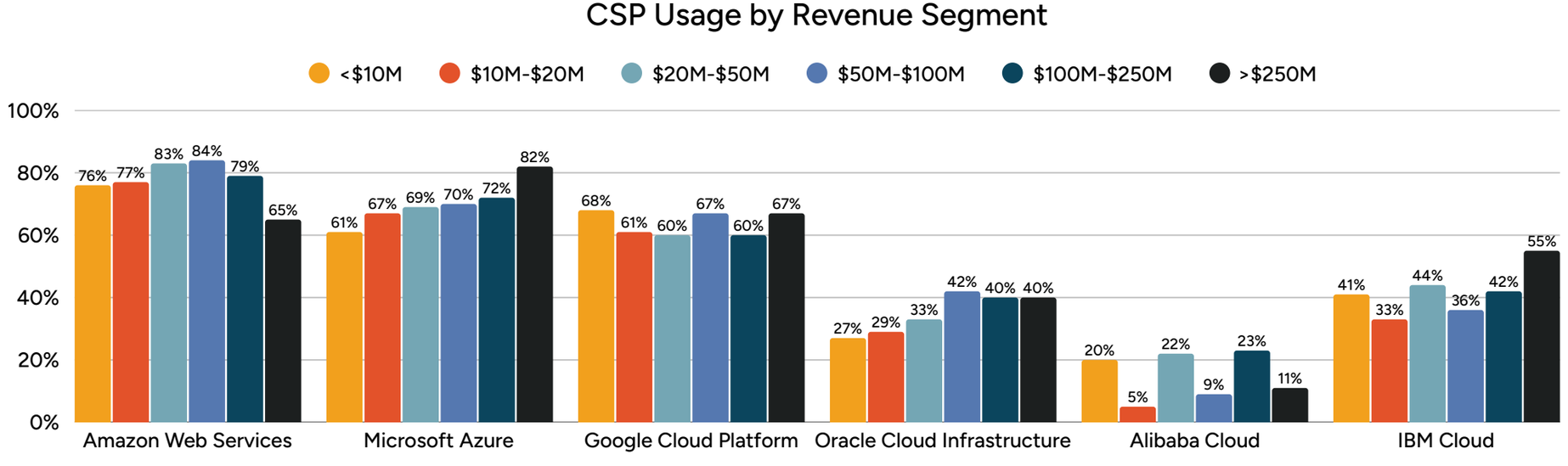
The architecture powering AI workloads is growing more complex, blending public clouds, private environments, and specialized AI service providers. While this diversity creates flexibility, it also fragments visibility, increases billing complexity, and introduces a wider range of unpredictable costs.

The research shows that hybrid is now the dominant model, multi-cloud usage is standard, and Microsoft Azure is rapidly gaining ground in the enterprise. These infrastructure decisions have direct consequences for cost governance and the ability to protect margins.

AWS remains the most widely used cloud provider overall (77%), followed closely by Microsoft Azure (71%) and Google Cloud (65%).

But the dynamic changes in the enterprise segment: among companies with more than $250 million in revenue, Azure adoption jumps to 82%, overtaking AWS. IBM Cloud ranks fourth in usage by number of companies, with strong penetration in specific verticals.

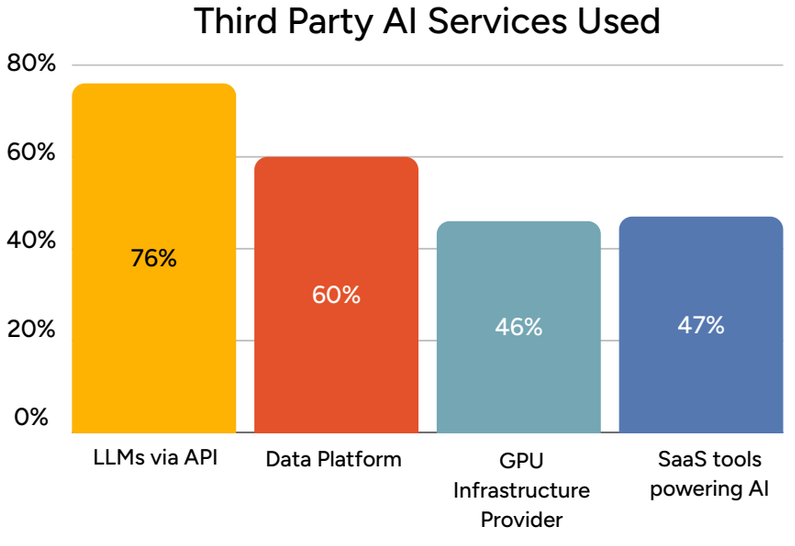
Most companies now operate in multi-cloud environments — often leveraging different providers for specific workloads, performance characteristics, or geographic needs. For CFOs, this means more contracts, more invoices, and more opportunities for spend to escape traditional oversight.

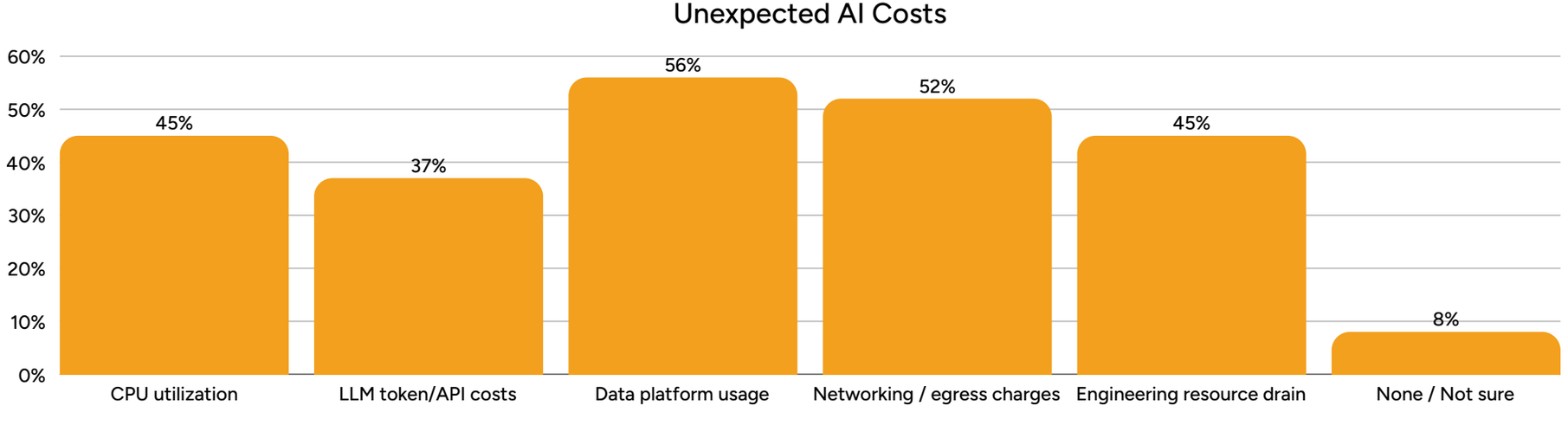


Third-party AI services add capability and cost diversity

Third-party large language models (LLMs) are the most common AI service (76% adoption). Data platforms, such as Datadog, are the second most common (60%) and they are the #1 source of unexpected AI costs. GPU infrastructure providers, such as CoreWeave, are used by 46% of companies.

Even companies that do not charge for AI-enabled products are heavy users of third-party LLMs (73%), meaning token-based costs are quietly reducing gross margins without being offset by direct revenue. For CFOs, this is a prime example of "hidden COGS" — costs that are real but unaccounted for in profitability models.

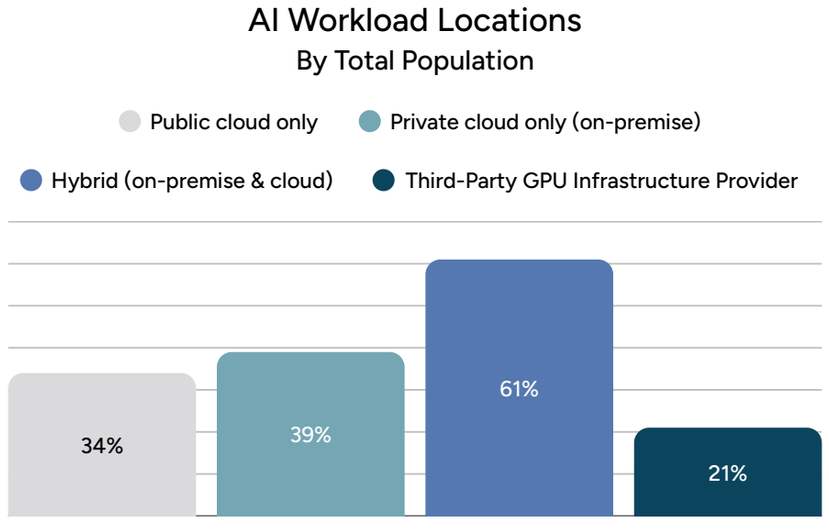




Hybrid complexity dominates workload placement

Hybrid, running workloads across both public and private cloud, is the most common model, used by 61% of companies. Only 34% run entirely in public cloud, and 21% use third-party GPU providers.

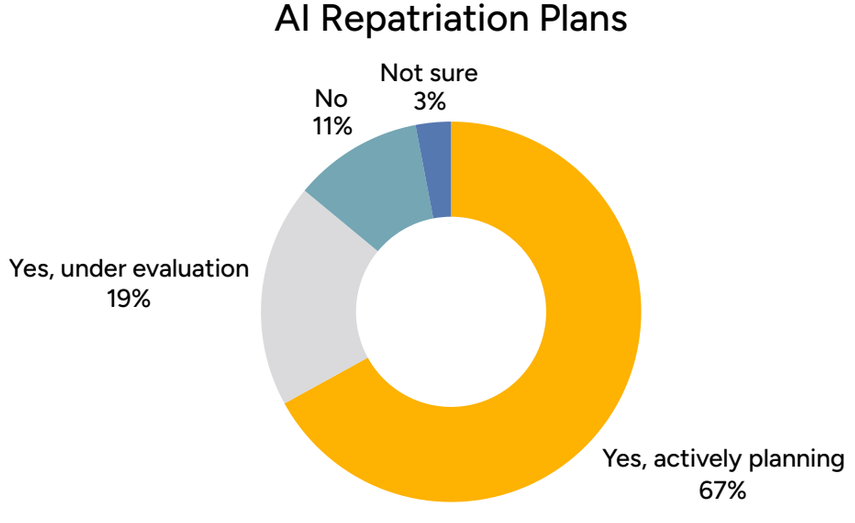
Hybrid is not just an enterprise pattern. Smaller companies (<$10M) show a 44% split between public and private cloud, proving that hybrid complexity can start early. From a financial perspective, hybrid environments often come with the highest cost visibility challenges, especially when cost reporting between public and private systems isn't standardized.



Repatriation is moving from plan to practice

Cloud-based training of large AI models can be prohibitively expensive at scale, with additional challenges around security, control, and performance consistency.

67% of companies are actively planning to repatriate at least some AI workloads to owned infrastructure, and another 19% are evaluating the move. Mid-market companies ($10M–$250M) show the highest planning rates, while large enterprises are more often in the evaluation phase. Companies that charge for AI are more likely to plan repatriation, linking monetization with tighter control over infrastructure.



CFO Takeaways

- **Infrastructure decisions have direct margin impact.** Hybrid and multi-cloud choices increase flexibility but multiply billing and visibility challenges.

- **The AI cost base is diverse and growing.** Data platforms, network access, GPU rentals, and LLM tokens each require their own forecasting models.

- **Repatriation is a financial strategy, not just a technical one.** It's a deliberate move to reshape the cost structure of AI workloads.

AI costs aren't just infrastructure, they're business risks hiding in your margins.

02 AI Cost Governance

Practices & Processes

For CFOs, the most pressing AI cost questions are rarely "How much did we spend?" Instead, we hear: "Do we know where the money went?" and "Can we predict what's coming next?"

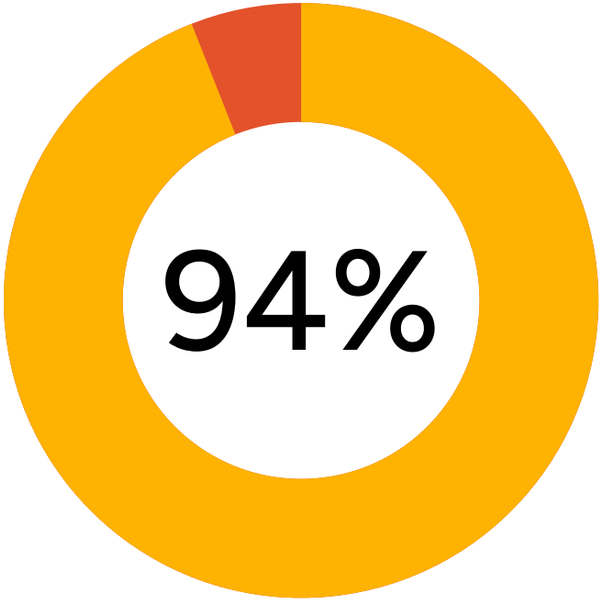
Our research shows that while most companies track some form of AI costs, gaps in process discipline, maturity, and visibility are keeping finance and product leaders from making fully informed decisions.

Charging for AI consistently correlates with stronger governance, yet a large segment of the market still delivers AI features "for free," and with far less cost control.

Tracking is common, but depth and timing vary

94% of companies say they track AI infrastructure costs, but the scope and granularity of that tracking differ widely. The key question is not if costs are tracked, but what's included, how early signals are captured, and who is accountable for acting on them.

Even among large enterprises (> $250M revenue), 3% admit they do not track AI infrastructure costs at all, surprising given the scale of spend. Smaller companies (<$10M) track at high rates (90%) but may lack the systems to measure at the same level of detail as larger peers.



of companies track AI infrastructure costs, but few capture them early enought to prevent budget surprises

Top challenges: visibility, forecast accuracy, hybrid complexity

When asked for their top three challenges in managing AI infrastructure costs, respondents most often cited:

1. Lack of visibility into costs (34%)
2. Inaccurate cost forecasts (16%)
3. Difficulty managing hybrid cloud environments (13%)

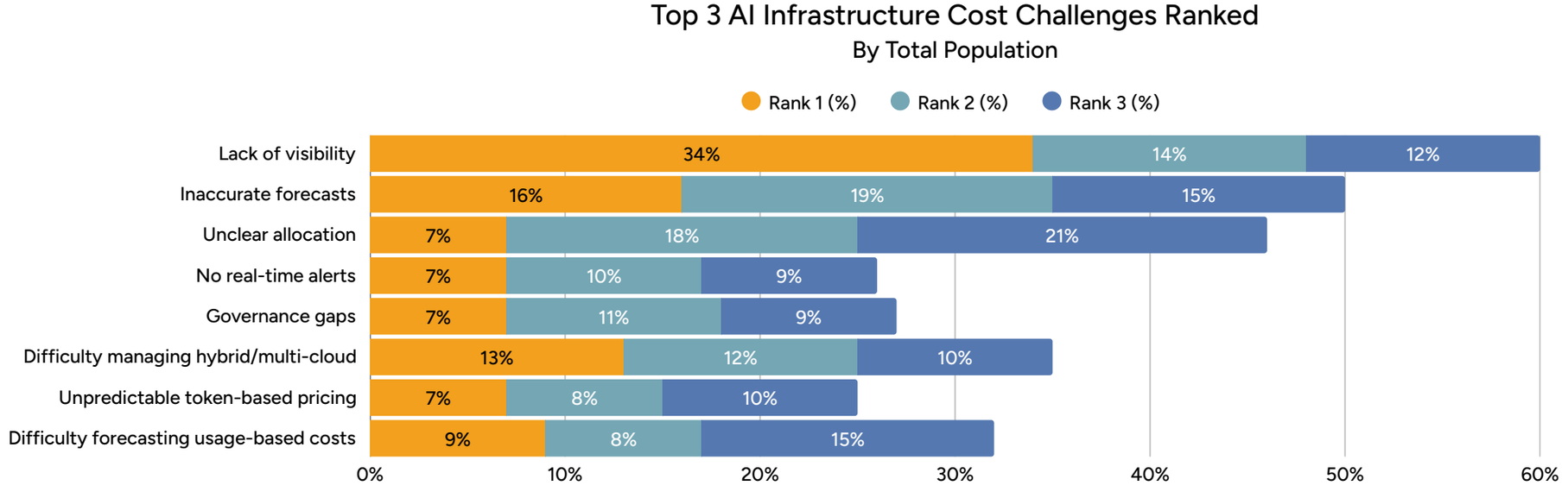
For finance teams, these challenges translate directly into higher gross margin risk and volatile forecasts.

Tactics for improvement

The most common tactic cited for improving AI cost management is unified visibility (33%): a single integrated view across all environments, services, and data pipelines. Clear cost attribution ranked second (22%), followed by better collaboration between teams (17%) and improved forecasting tools (15%).

Budgets exist, but don't guarantee control

94% of companies that track AI costs also assign an AI infrastructure budget, though budgeting is slightly less common in the smallest companies (88%).



The Revenue Accountability Effect

Why Charging for AI Drives Better Governance

One of the most striking patterns in our research isn't about technology, it's about incentives. Companies that charge for AI consistently demonstrate superior cost discipline across every metric we measured. This isn't coincidence; it's the power of revenue accountability.

Consider the stark differences:

- **70% of companies charging for AI** can track cost-to-serve precisely, compared to just **29% of those giving AI away**
- **71% use real-time usage alerts** for overages, versus much lower rates among free providers
- They're **twice as likely** to attribute costs by customer, product, or AI model
- They're **significantly more likely** to include AI costs in strategic decision-making

Why does charging create this discipline? Three forces are at work

1. **Customer Pressure Creates Operational Rigor:** When customers pay for AI features, they expect value and reliability. This external pressure forces internal teams to understand exactly what they're delivering and what it costs. Every support ticket about slow performance or unexpected charges becomes a forcing function for better cost attribution.

2. **P&L Ownership Changes Behavior:** When AI moves from a cost center to a profit center, someone's bonus depends on managing those margins. Product managers start asking "What's our cost per inference?" Finance teams demand real-time dashboards. Engineering teams optimize for efficiency, not just functionality.

3. **Pricing Decisions Require Cost Truth**: You can't price what you can't measure. Companies charging for AI are forced to develop granular cost models to stay competitive and profitable. This requirement drives investment in the very systems that enable better governance.

The Hidden Cost of "Free" AI: Meanwhile, companies providing AI features at no charge often treat AI costs as overhead—a dangerous blind spot. Without revenue pressure, these costs can balloon unchecked.

The Governance Paradox: Interestingly, companies charging for AI also show the highest rates of early-stage cost management maturity (34%), suggesting that monetization triggers governance investment, even if many are still building the foundation. It's proof that revenue accountability accelerates learning, even when systems aren't perfect.

For CFOs, the implication is clear: if you're giving AI away for free, you're not just missing revenue, you're missing the accountability mechanisms that drive cost control.

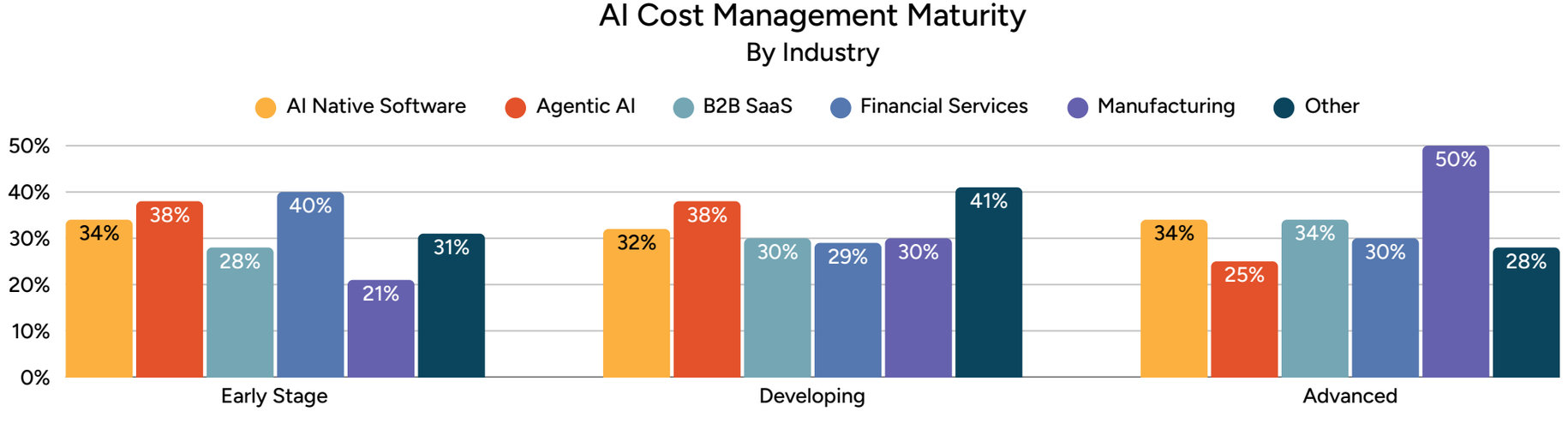
Maturity Levels

Most companies are still in early or developing stages

Only 34% of companies say they have an "advanced" AI cost management program, defined as having tracking, cost attribution, and governance policy in place.

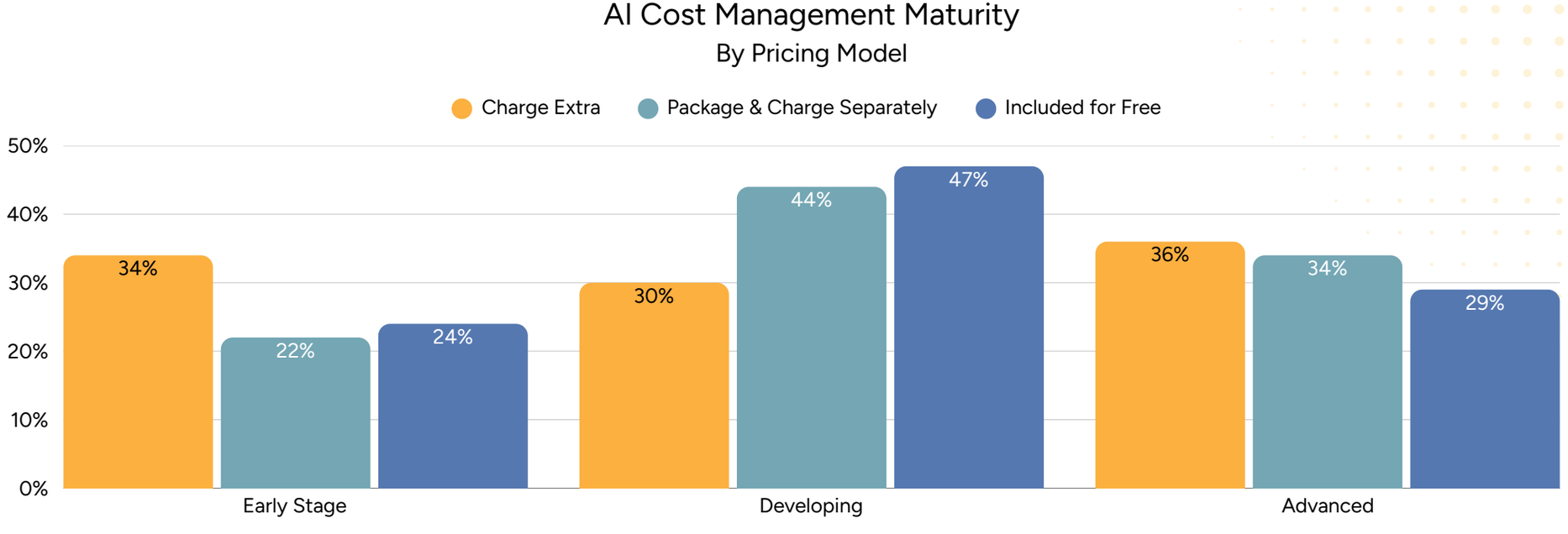
- **Early stage:** 30% are just starting to track, budget, and allocate costs
- **Developing:** 36% have some visibility, but mostly manual processes
- **Advanced:** 34% have automated tracking and attribution with governance in place

Industry plays a bigger role in maturity than company size. Manufacturing leads (50% advanced), while Financial Services (40% early stage) and Agentic AI companies (38% early stage) lag.



Charging for AI also correlates with higher maturity

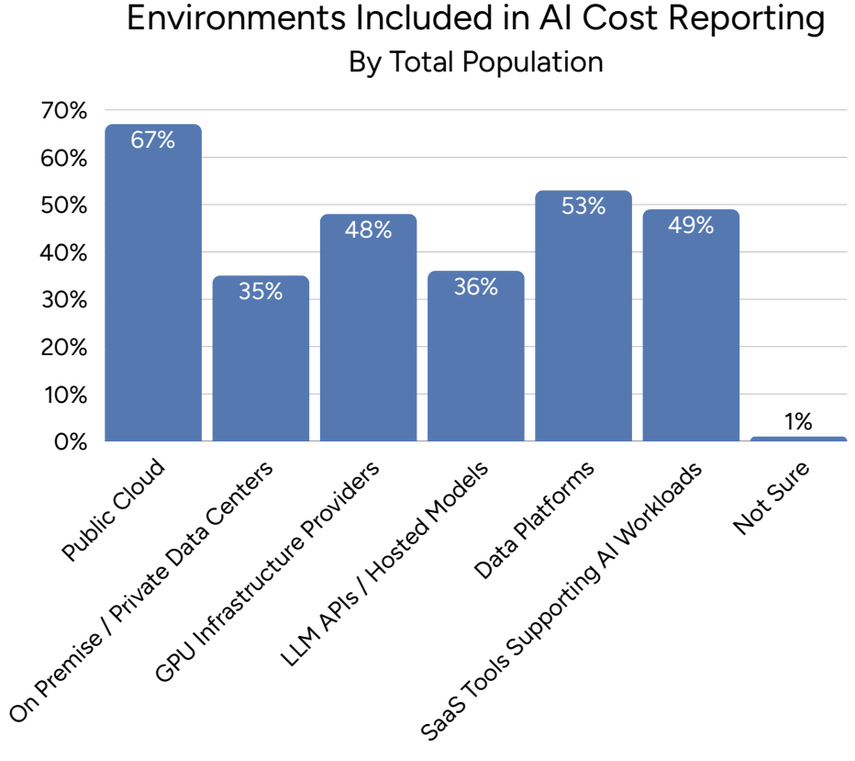
Companies that charge for AI products or features are more likely to be advanced in maturity (34–36%). Surprisingly, they also show the highest percentage of early-stage maturity (34%), suggesting that monetization triggers governance investment, although many are still building the foundation.



Visibility & Attribution

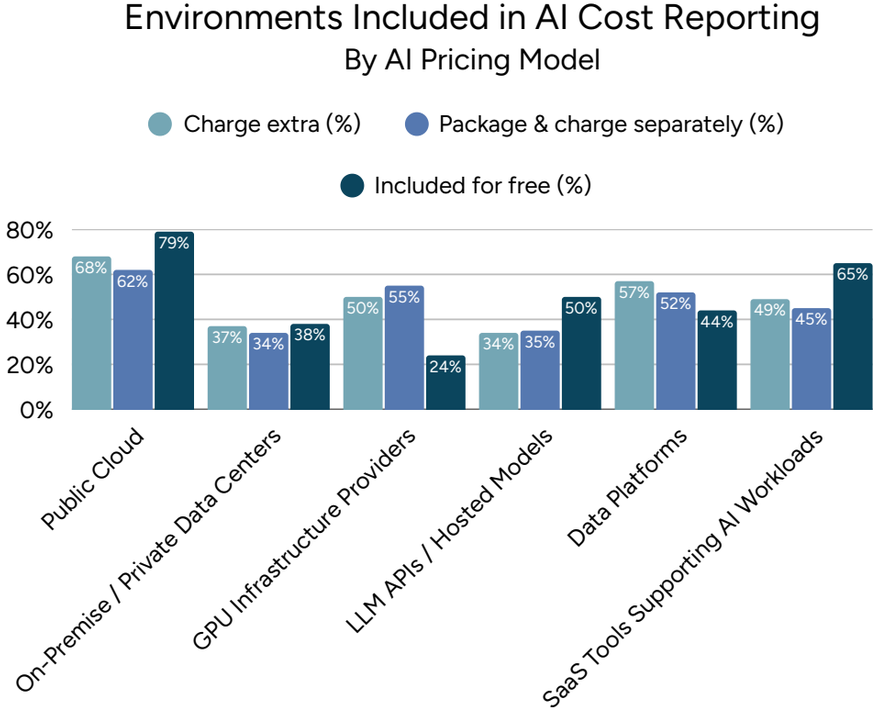
On-prem costs are often invisible

Only 35% of companies include on-premise AI infrastructure in their cost reporting. Cloud and third-party providers typically offer better native reporting, but gaps in on-prem data create major blind spots, especially in hybrid models.



LLM API costs not always included, even when core to product

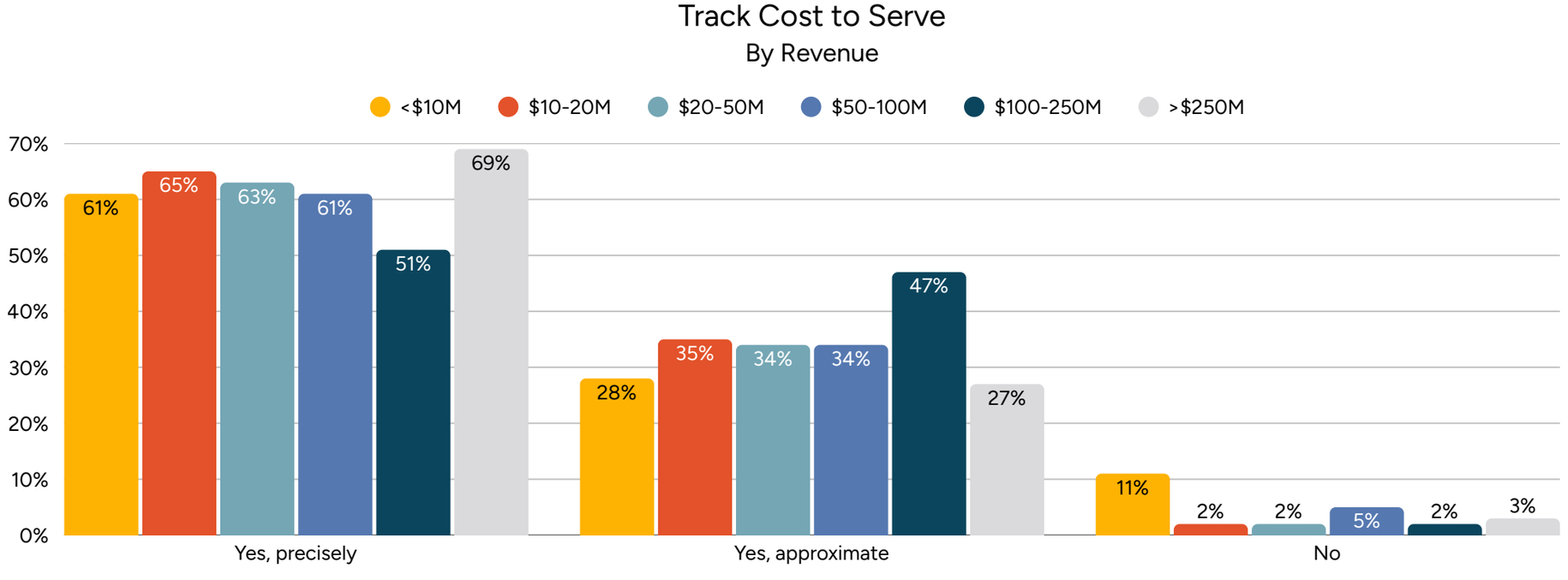
Only ~50% of companies using AI as a core part of their product include LLM API costs in their AI cost reporting. This omission makes true cost-to-serve and gross margin calculations unreliable.



Precise cost-to-serve tracking is far from universal

While 62% of companies can track cost-to-serve precisely, 34% can only track approximately, and a small segment (4%) cannot track it at all. Companies that charge for AI features or package AI as a separate paid product and AI-native providers are far more likely to have precision tracking in place.

Precise cost-to-serve tracking separates margin leaders from laggards.

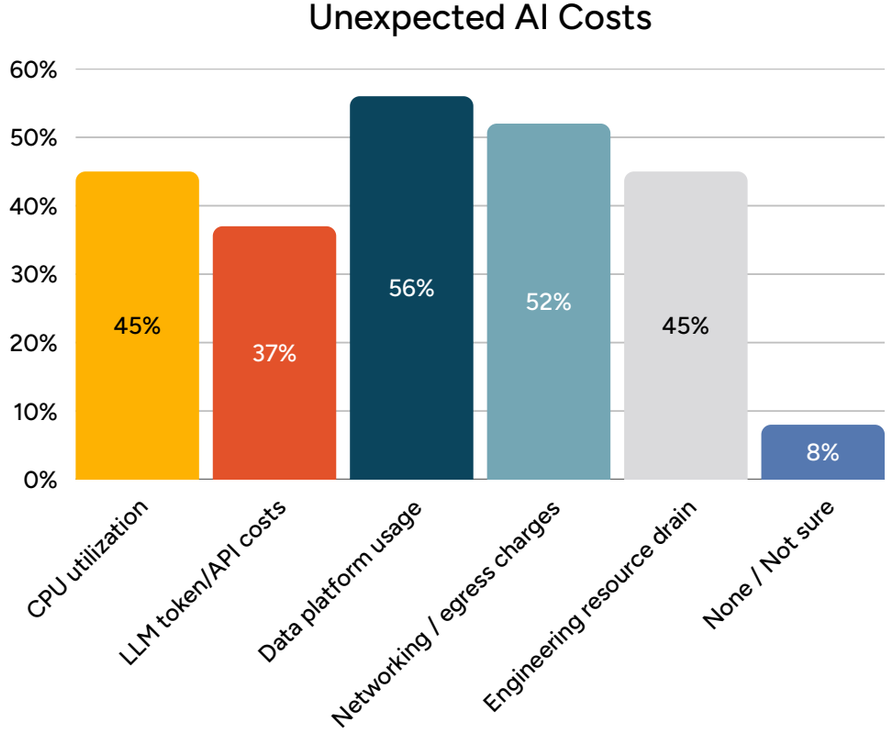


Unexpected costs go beyond tokens

The top two unexpected cost drivers are:

1. Data platform usage (56%)
2. Network access to AI models (52%)

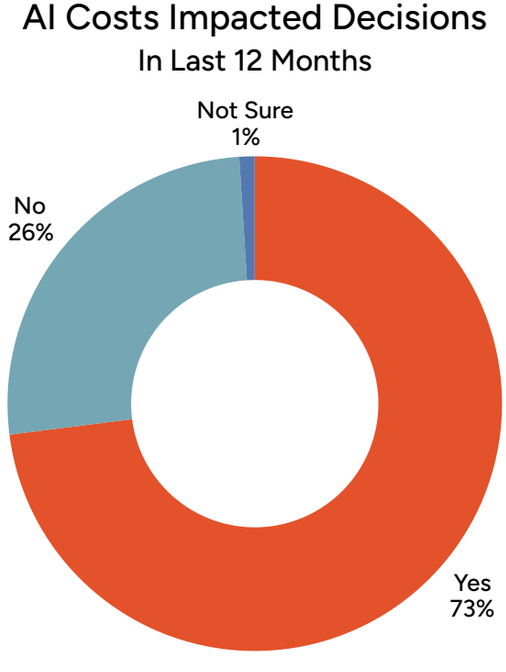
LLM token costs, often assumed to be the main culprit, rank only fifth (37%).



Decision-making suffers without full visibility

26% of companies say AI costs did not impact any major decisions in the past year: a sign that cost data isn't being used strategically.

Companies that charge for AI are far more likely to factor costs into pricing, packaging, and infrastructure decisions.



CFO Takeaways

- **Budgets exist, but attribution lags.** 94% of respondents assign AI budgets, yet only 35% include on-prem costs and just half report LLM API usage, leaving major blind spots.

- **Governance maturity varies by industry, not size.** Sectors like financial services are further ahead, while others struggle to move beyond basic cost tracking.

- **Monetization drives discipline.** Companies that charge for AI are more likely to track cost-to-serve precisely, apply real-time usage alerts, and attribute costs by product or model.

You can't govern what you can't see; and without full visibility and attribution, even the best-intentioned AI budgets are leaving worrisome blind spots.

03 AI Costs: Financial Management & Metrics

Measuring Financial Impact

For CFOs, the numbers tell the story

AI is no longer an experimental budget line: it's a material cost driver affecting gross margins, profitability, and financial predictability.

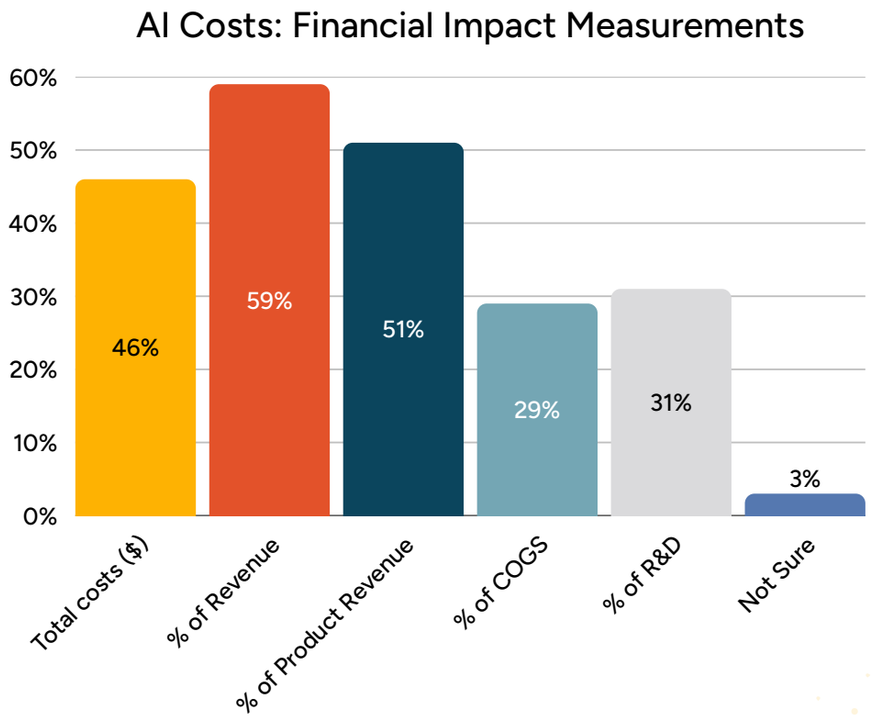
This section captures how companies measure AI's financial impact, how accurately they forecast spend, and what tools they rely on to manage usage.

Most companies measure AI costs as a percent of revenue

59% of companies measure AI infrastructure costs as a percentage of revenue.

While this is aligned with common cloud cost reporting, **only 29% measure AI costs against COGS, which is the metric most closely tied to gross profit.**

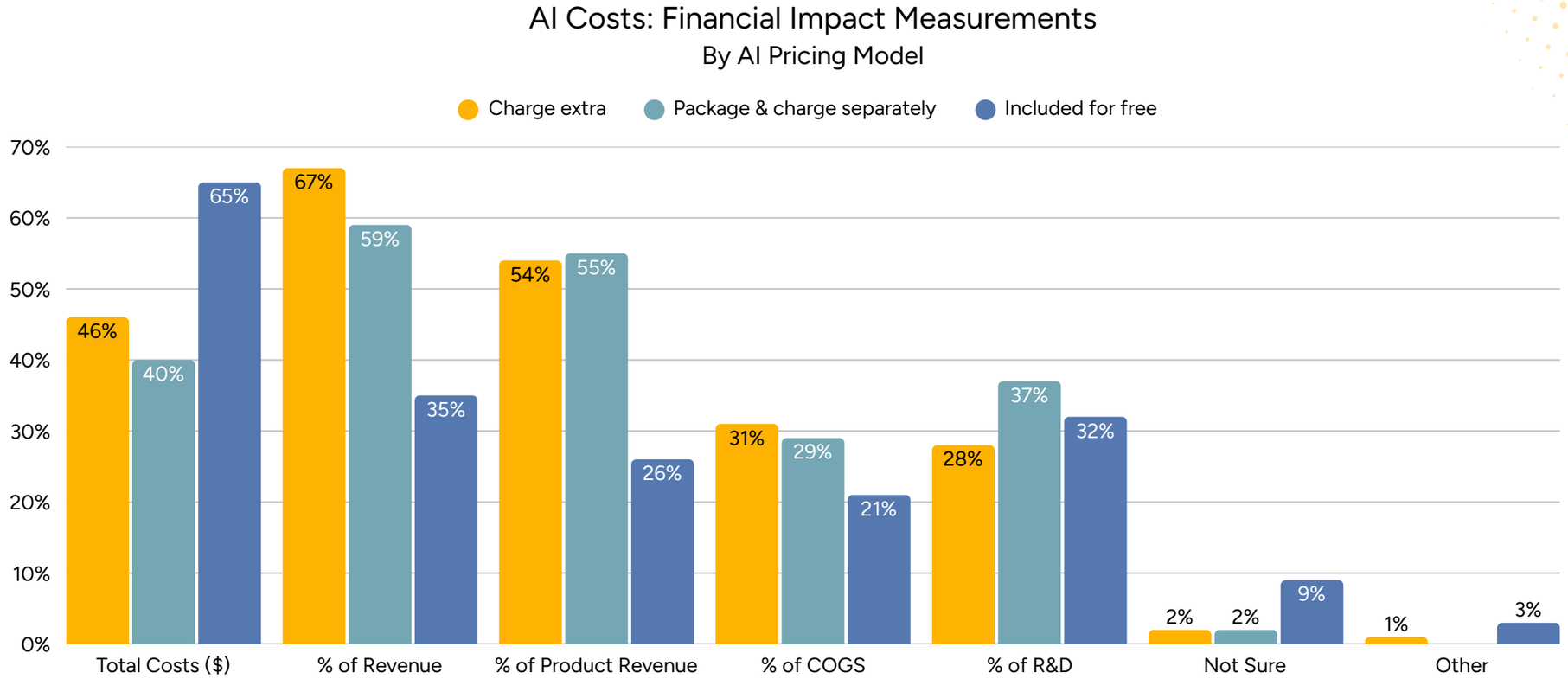
A smaller group measures AI costs as a percentage of R&D, which may signal a shift in how R&D is treated on the income statement.



Charging for AI sharpens profitability tracking

Companies that package or charge for AI products are more likely to measure costs against total revenue, product-specific revenue, and COGS.

Those providing AI for free risk missing key profitability signals.

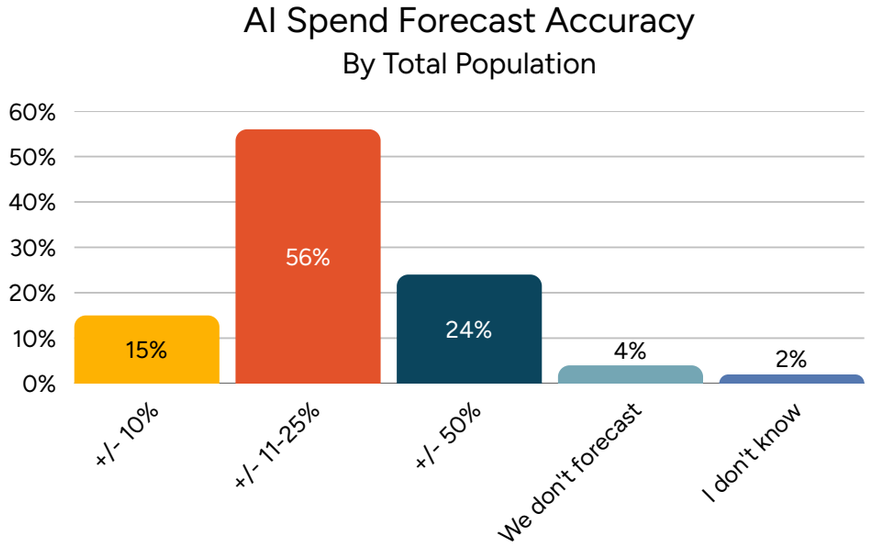


Forecast Accuracy

Forecast misses are widespread

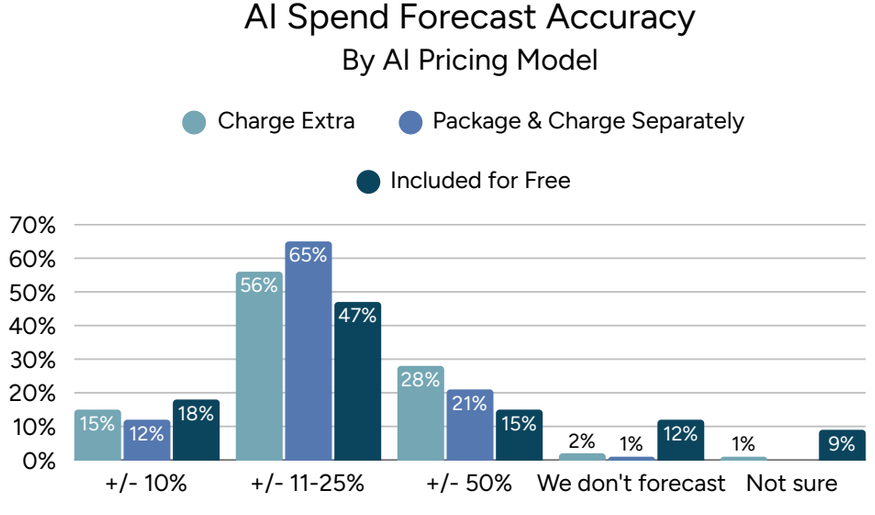
Only 15% of companies forecast AI costs within ±10%. A majority, 56%, miss by 11–25%, and nearly one in four (24%) miss by more than 50%.

This level of inaccuracy puts gross profit targets at risk, especially as AI costs become a larger share of COGS.



Charging for AI does not guarantee accuracy

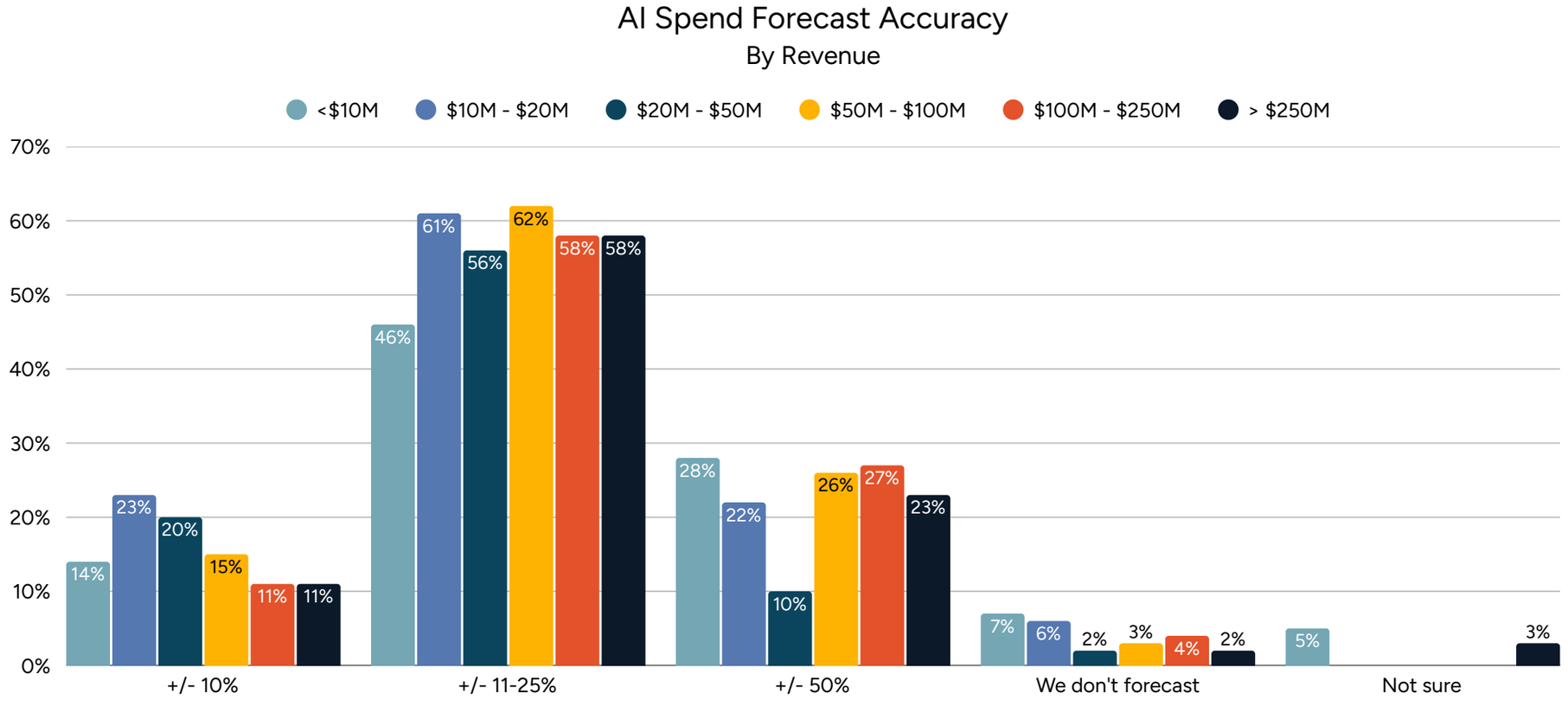
Interestingly, companies that charge for AI are more likely to have the largest forecast misses (>50%) compared to those that do not. This points to a gap between monetization strategy and operational forecasting capability.



Forecast accuracy is not tied to company size

Large enterprises are just as likely to miss forecasts by wide margins as small companies.

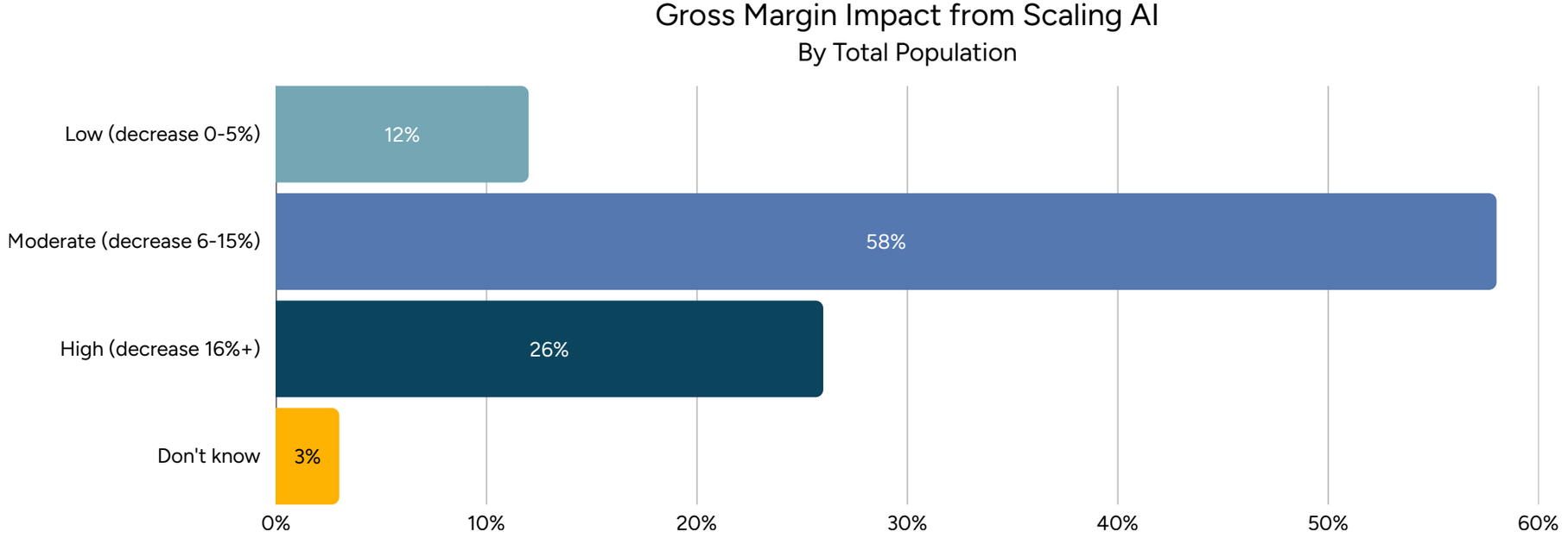
Companies in the $10M–$50M range are the most likely to hit within ±10%, possibly due to being AI-native and having built granular cost tracking earlier.



Gross Margin Impact

Most companies see margins fall by 6% or more

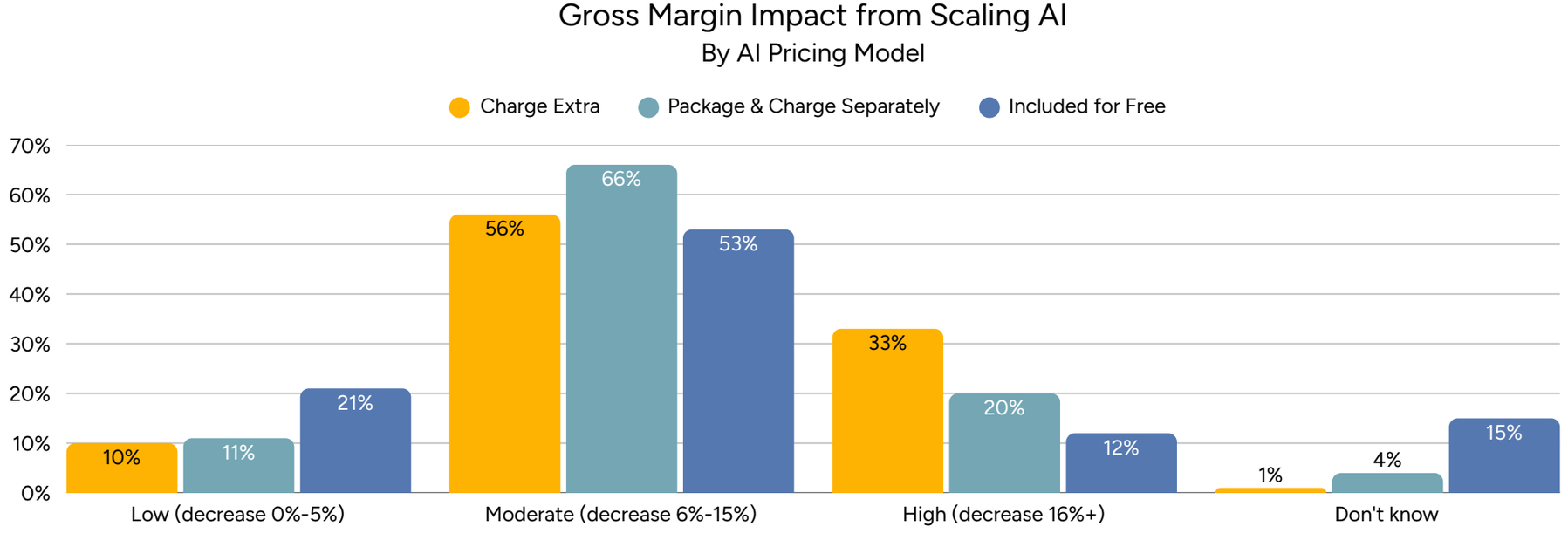
84% of respondents report more than a 6 percentage point (600 bps) drop in product gross margin due to AI delivery costs. For example, a product at 80% gross margin could fall to 74%. For 58% the impact is 6–15 points (600–1500 bps), and for 26% it is 16+ points (1600+ bps). This level of compression requires active cost governance to maintain profitability targets.



Monetized AI products see the largest hits

Many companies still detect AI usage overages only after receiving invoices. The largest enterprises are the most likely to rely on manual reviews, a lag that increases the risk of runaway costs.

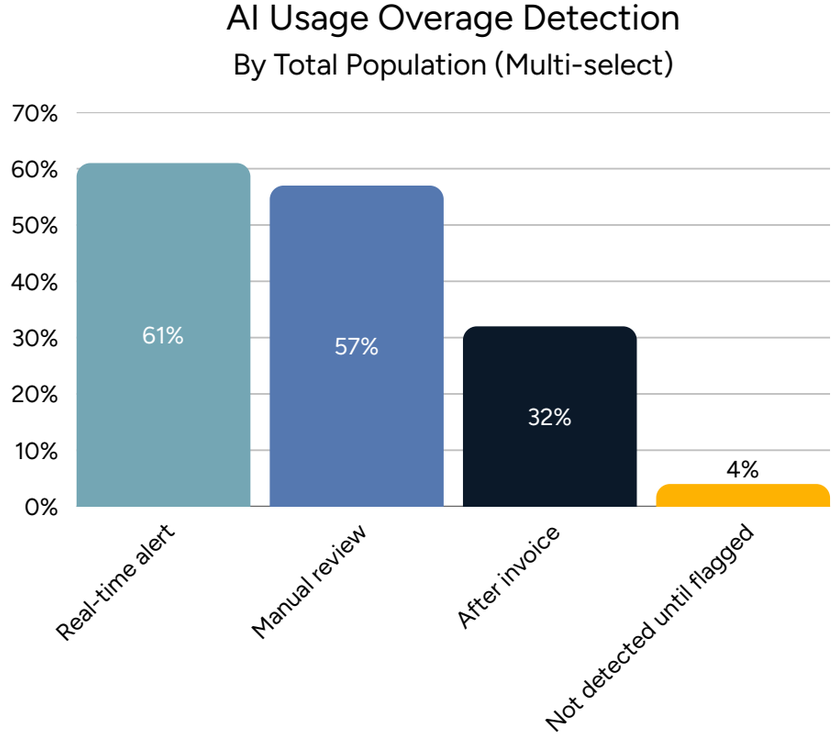
When a third of companies charging for AI see margin hits above 16%, it's no longer just an infrastructure problem, it's a pricing, packaging, and profitability issue.



Usage Overage Detection

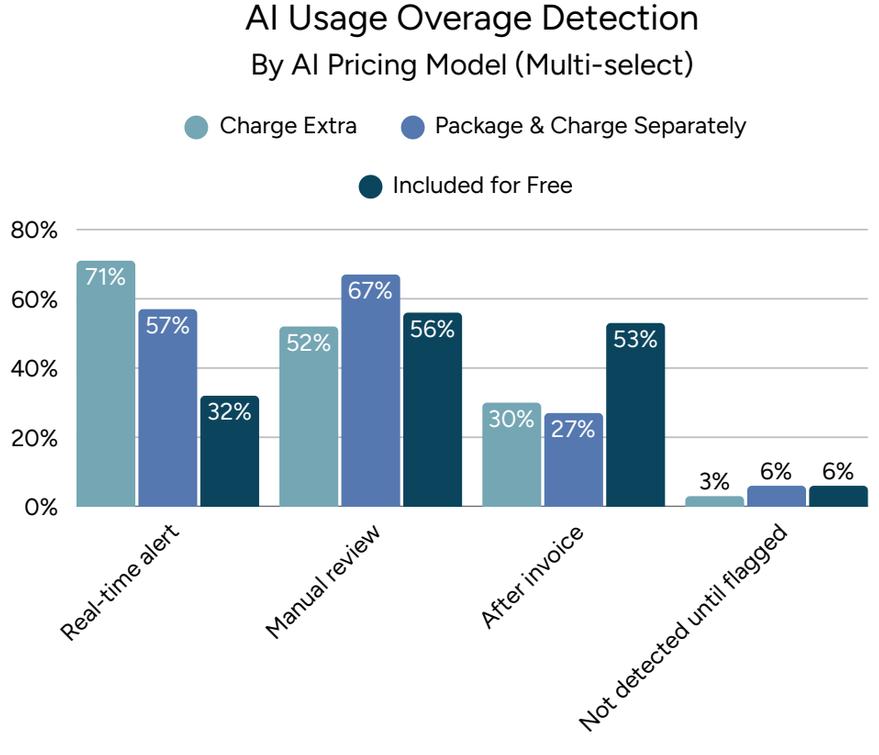
Real-time monitoring is not universal

Many companies still detect AI usage overages only after receiving invoices. The largest enterprises are the most likely to rely on manual reviews, a lag that increases the risk of runaway costs.



Charging companies are more likely to use alerts

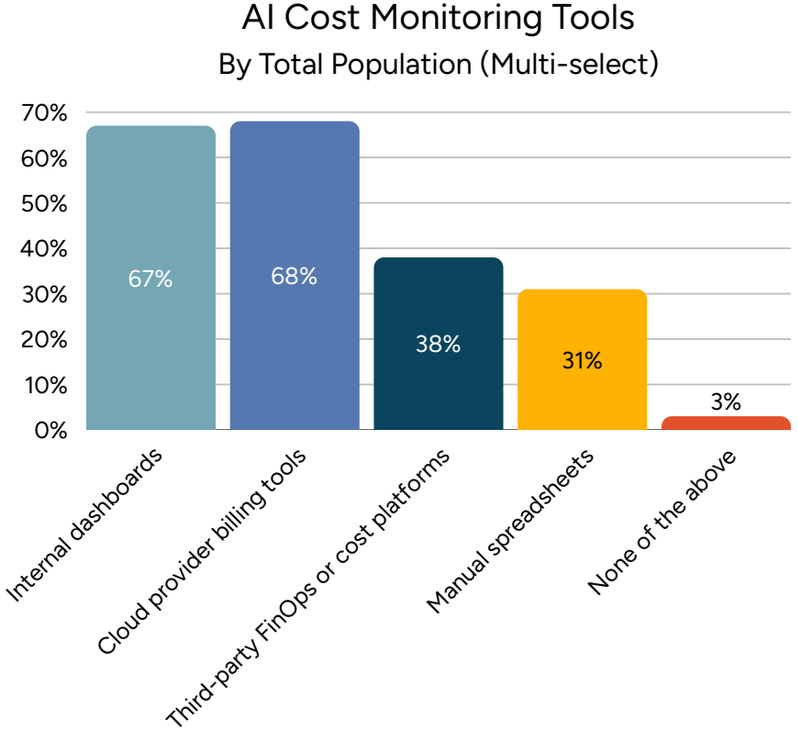
71% of companies that charge for AI use real-time alerts for overages compared to much lower rates among those who give AI away.



Monitoring Tools

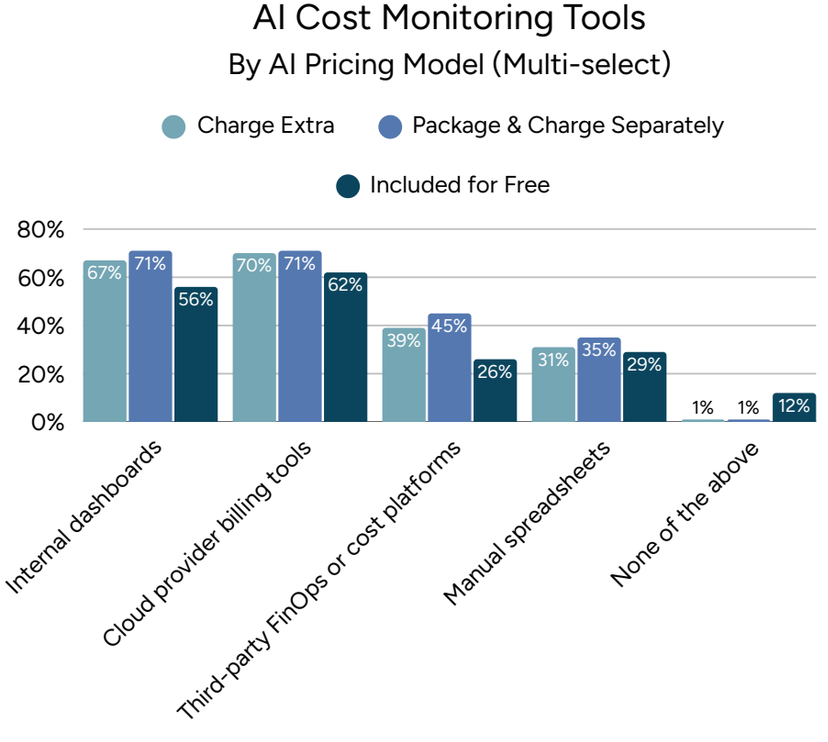
CSP tools and internal dashboards dominate

The most common monitoring methods are cloud service provider tools (68%) and internally developed dashboards (67%). Only 38% use specialized third-party cost management platforms, raising questions about whether current tools deliver early enough warning signals.



Companies that charge for AI offerings ultimately invest more in purpose-built tools

Those building and monetizing AI products are more likely to use a combination of internal dashboards, CSP tools, and dedicated cost management platforms.



CFO Takeaways

- **AI forecast misses are the rule, not the exception.** 85% of respondents miss AI cost forecasts by more than 10%, and nearly one in four miss by 50% or more.

- **Gross margin erosion is already material.** 84% of respondents report AI delivery costs cutting product gross margins by more than 6 points (600 bps), with 33% of those who charge for AI seeing hits of 16+ points (1600+ bps).

- **Monitoring tools remain basic.** Most companies rely on native CSP tools and disjointed internal dashboards, rather than unified cost management.

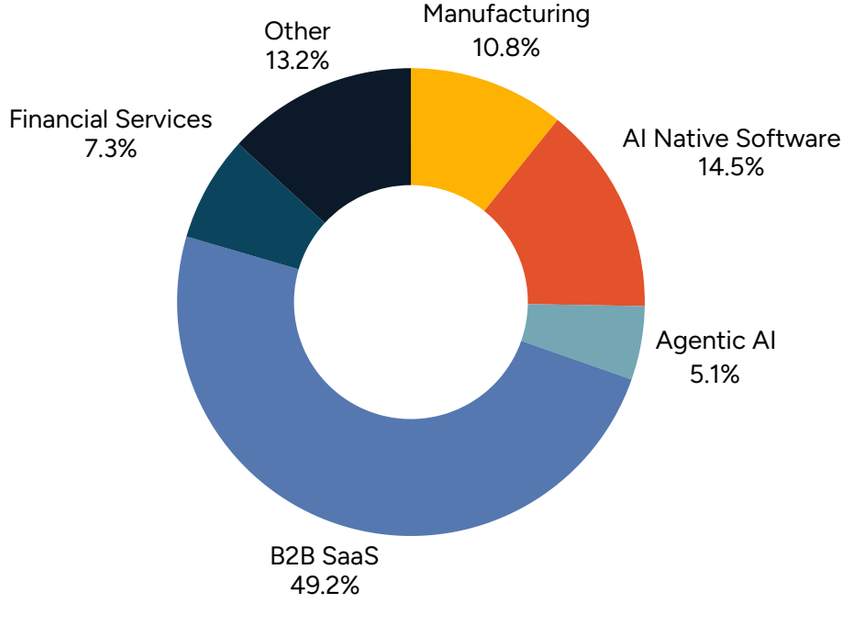
AI spend isn't just hard to predict, it's actively reshaping gross margins, and most companies are still managing it with little more than rear-view mirrors.

04 Participant Profile

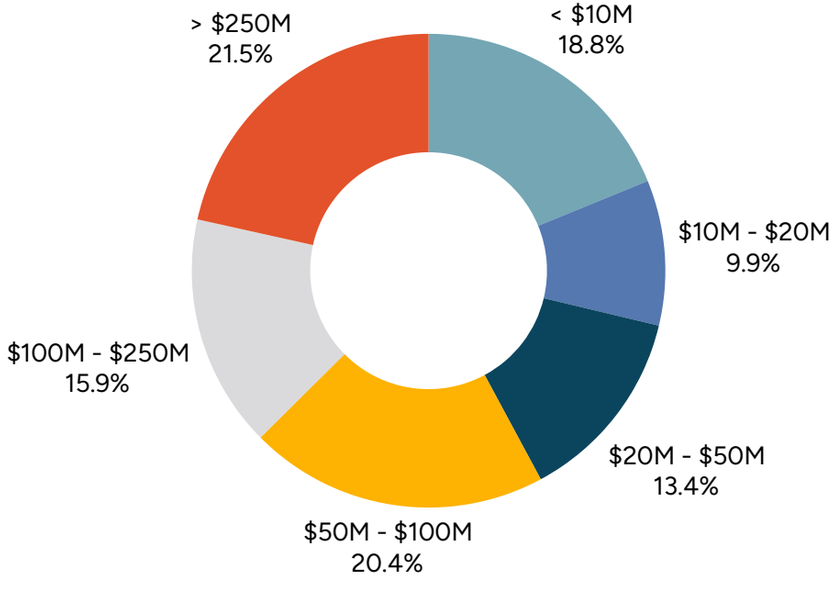
Participant Profile

To interpret the benchmarks in this report, it's important to understand who participated in the research. The 372 respondents represent a diverse mix of industries, company sizes, AI maturity levels, and go-to-market models.

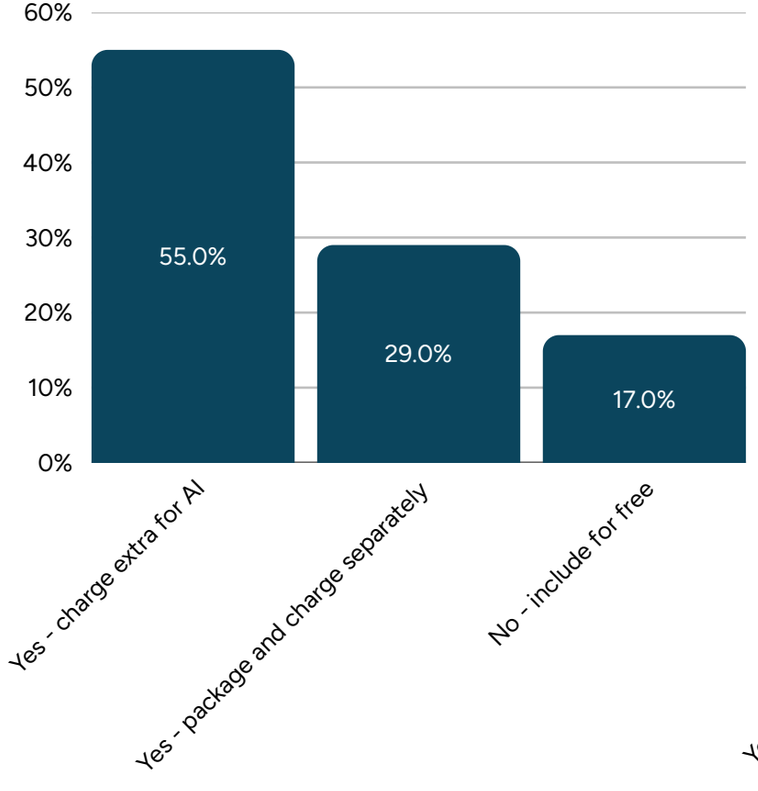
Industry Distribution



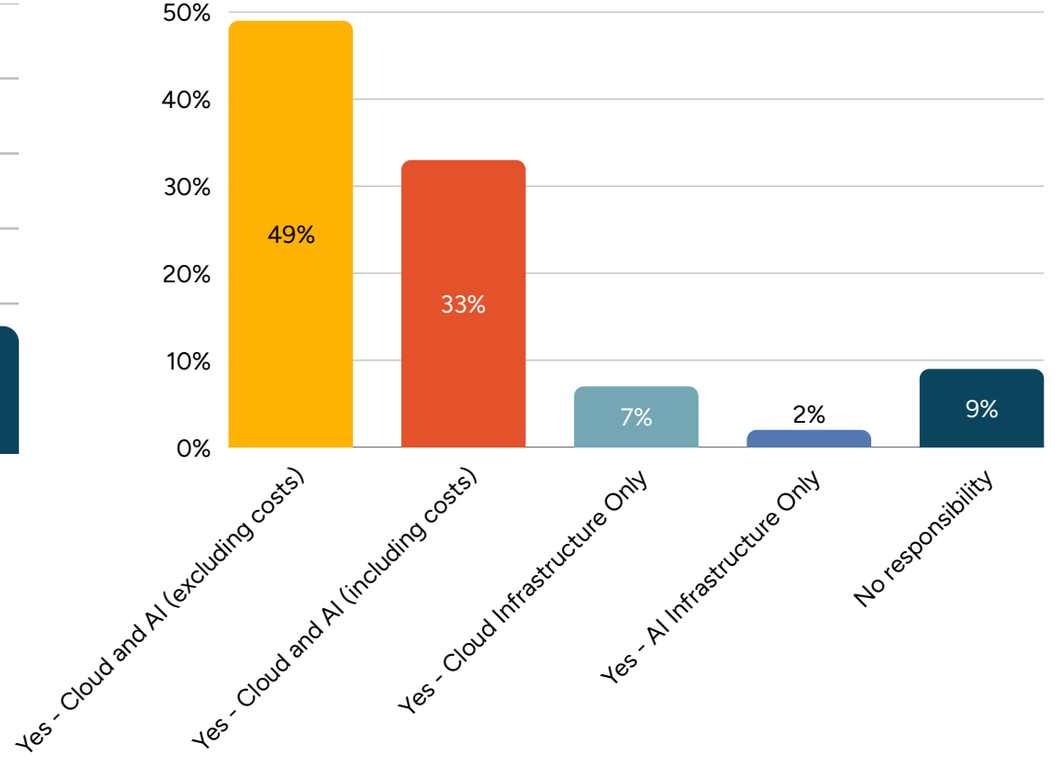
Company Size by Revenue



AI Adoption Models



AI Responsibility



Conclusions: Preparing for 2026

AI is no longer an experimental line item. It's a structural cost with direct impact on margins, forecasts, and shareholder value. As we move into 2026, the companies that thrive will be those that treat AI cost governance not as an afterthought, but as a core pillar.

Key Considerations for CFOs & Finance Leaders

- **Forecasting will separate leaders from laggards.** With 85% of organizations missing AI cost forecasts by more than 10%, the winners in 2026 will be those who embed forecasting discipline into every AI initiative.
- **Margin pressure isn't optional anymore.** AI's financial impact is already material. Gross margin reductions of 6–15 percentage points (600–1500 bps) are common, and over a quarter of companies see hits of 16+ points (1600+ bps). CFOs must treat AI costs as part of COGS, not just "innovation expense."
- **Visibility is the foundation of governance.** On-prem blind spots, hidden LLM API usage, and fragmented billing create governance risk. As AI spend grows, financial control requires unified reporting across every environment.
- **Hybrid has a new definition.** It's no longer just public + private cloud. Today, hybrid means blending cloud with AI on-prem; repatriated training clusters, owned GPU infrastructure, and local data platforms, all reshaping cost models and governance requirements.

mavvrik.

Mavvrik is the financial control center for modern IT, providing enterprises with complete visibility, automation, and governance across cloud, AI, SaaS, and on-prem infrastructure. Built for CFOs, FinOps, and IT leaders, Mavvrik eliminates financial blind spots, enforces accountability, and transforms IT costs into strategic investments. With real-time cost tracking, automated chargebacks, and predictive budget controls, Mavvrik helps enterprises reduce waste, optimize AI and hybrid cloud spend, and maintain financial precision at scale.

For more information visit www.mavvrik.ai.

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 www.benchmarkit.ai