

July 2026
InfoBrief
Sponsored by Microsoft

Scaling AI Without Scaling Costs

Best Practices from 3,000 Organizations Building Financial Confidence into AI



Kathy Lange
Research Director,
AI Software, IDC

Table of contents

Click any title to navigate directly to that page.

In this InfoBrief 3

Key findings: From AI ambition to economic reality 4

Planning AI adoption: AI funding has moved beyond experimentation 5

AI financial planning is formalized but often late in the process 6

Estimating the cost of AI: Confidence is high but uneven 7

AI budget forecasting challenges diverge by scale 8

Productivity is the top priority, but ROI and cost predictability are the gates 9

Designing AI for efficiency: Early choices shape long-term costs 10

Allocation of AI budgets in the design and build stage 11

AI unit economics awareness is emerging as a core design practice 12

Cost-efficient models don't guarantee cost-efficient AI 13

Most organizations prioritize AI application quality over cost 14

Managing AI investments: AI is controlled, but learning lags 15

When AI hits production, accountability is centralized 16

Financial measures tracked for AI solutions in production 17

Production insights do not consistently shape what gets built next 18

Measuring return on AI investment: Growing focus on demonstrating value 19

Linking AI spend to financial impact is still maturing 20

AI has become a CFO-level topic, not just a CIO project 21

Quantifying AI productivity gains distinguishes real AI value 22

AI operates under a defined ROI window 23

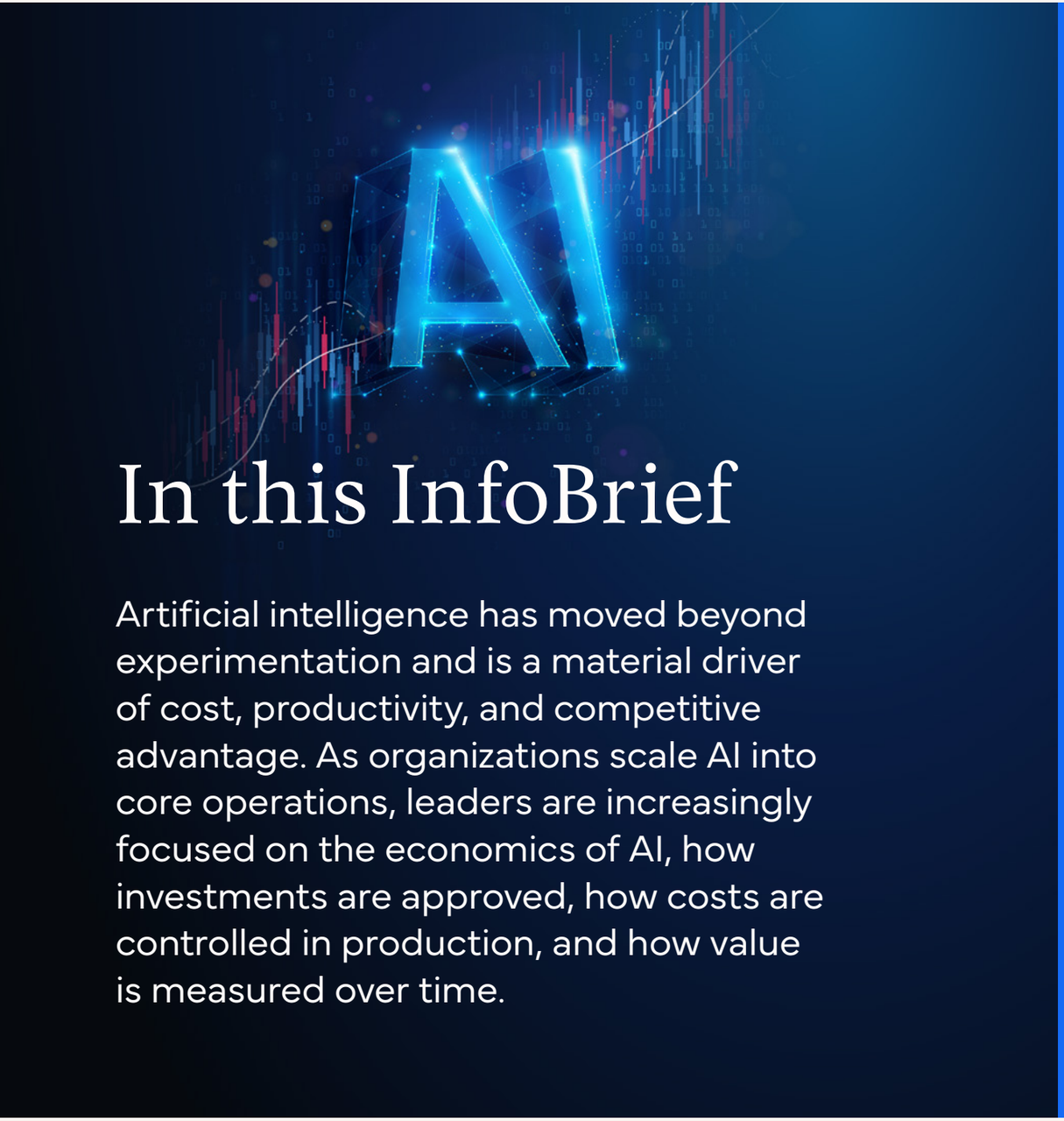
AI is delivering strategic value, with financial returns emerging 24

Strategic value dominates today, and financial returns are next 25

Guidance for managing AI as an economic investment 26

About the IDC analyst 27

Message from the sponsor 28



In this InfoBrief

Artificial intelligence has moved beyond experimentation and is a material driver of cost, productivity, and competitive advantage. As organizations scale AI into core operations, leaders are increasingly focused on the economics of AI, how investments are approved, how costs are controlled in production, and how value is measured over time.

Worldwide survey

To understand how organizations are navigating the economics of AI, IDC conducted a worldwide survey of 3,000 enterprises and small and medium businesses across industries to examine how they plan, design, manage, and measure AI initiatives from a financial perspective. **The research focused on best practices, financial thresholds, cost allocation, unit economics, production oversight, and the balance between short-term returns and long-term strategic value.**

This InfoBrief highlights the key findings from the study and shows how the economics of AI are evolving as adoption matures. It surfaces where organizations are gaining discipline and where gaps remain. Together, these insights offer leaders a clearer view of how to manage AI not just as an innovation effort but as a scalable, ongoing economic investment.

Note: All numbers in this document may not be exact due to rounding. All data from IDC's *Microsoft Custom Survey 2026: Financial Impact of AI;* n = 3,030

Key findings: From AI ambition to economic reality



AI spending is becoming more tightly managed as a financial investment, not an experiment.

Nearly 90% of organizations apply financial thresholds before approving AI initiatives, and 88.5% report centralized financial oversight, confirming AI has become a CFO-level priority.



Planning confidence is high, while managing cost variability remains a key focus area.

While almost 94% of organizations say they can estimate AI costs, nearly half cite the rapidly changing AI model landscape as their biggest budgeting challenge.



Cost discipline weakens after design decisions are made.

Although 77.7% prioritize cost efficiency in model selection, only 17% build dedicated models to project how AI costs scale in production, leaving long-term unit economics underdeveloped.



Production oversight exists, but learning rarely feeds planning.

Nearly half of organizations centralize AI cost monitoring in production, yet only about one-third have formal feedback loops that use production data to guide future AI investments, limiting compounding returns.



ROI expectations are defined, but measurement precision lags.

Over 86% expect AI to deliver financial results within 6–24 months, yet fewer than half say they can link AI initiatives to financial outcomes very well, highlighting an opportunity to strengthen value measurement as execution scrutiny increases.

Planning AI adoption:

AI funding has moved beyond experimentation

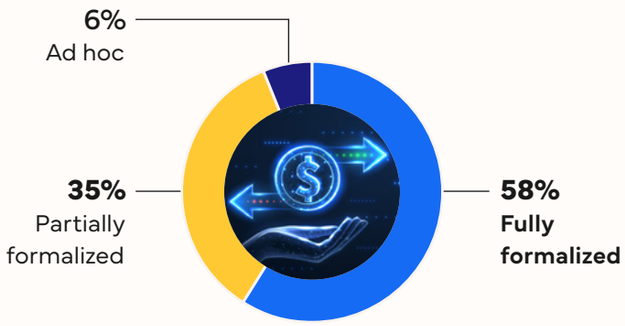
2 3 4

AI financial planning is formalized but often late in the process

The era of shadow AI budgets is ending.

Processes

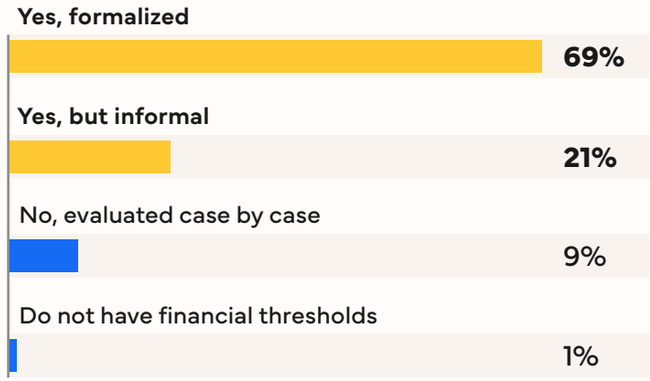
In the survey, 58% reported fully formalized financial planning processes with clear criteria, and another 35% have partially formalized processes.



Thresholds

A striking 90% of organizations have established financial thresholds before approving new AI initiatives (69% formally and 21% informally).

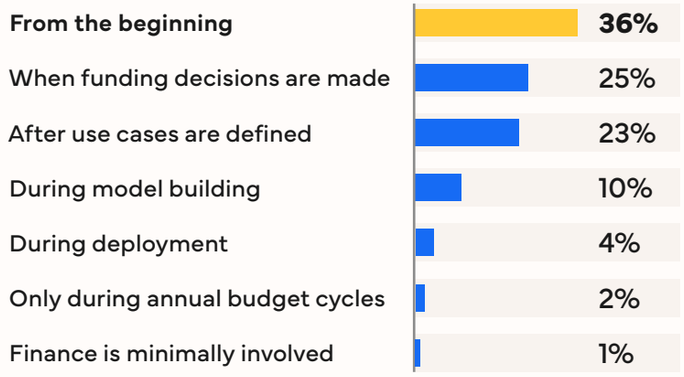
Does your organization set specific financial thresholds before approving AI initiatives?



Timing

But only 36% involve finance from the beginning; 39% wait until funding is already being decided or later, creating a planning gap where cost assumptions go unchallenged.

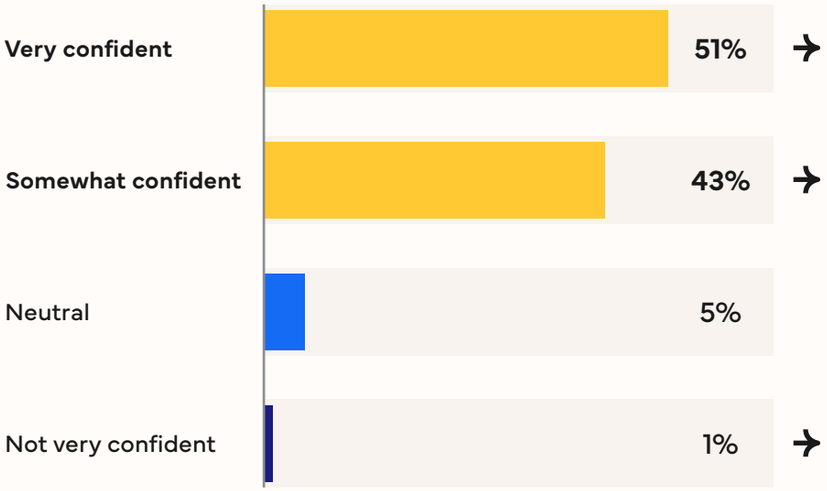
How early are finance teams involved in new AI initiative planning?



Estimating the cost of AI: Confidence is high but uneven

Nine out of ten organizations say they can accurately estimate AI costs. The gaps that remain in companies are where AI is the most complex.

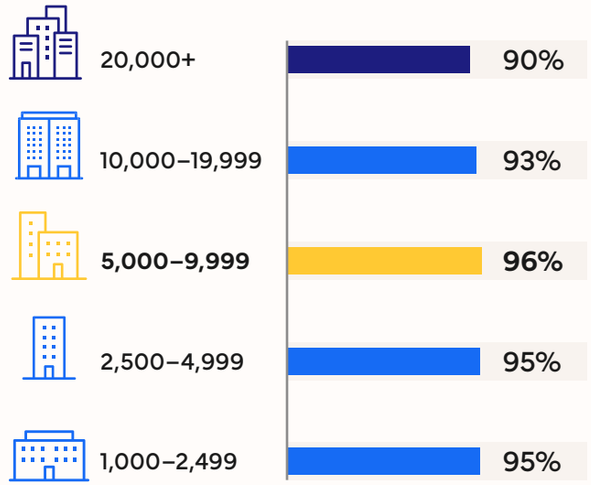
How confident are you in your organization's ability to accurately estimate AI initiative costs?



Companies with 5,000 to 10,000 employees are the most certain, with 94% expressing they are "very" or "somewhat" confident.

The largest organizations, those with 20,000 or more employees, are the least certain (90%), suggesting that as AI scales across a complex enterprise, cost estimation gets harder, not easier.

Enterprise size:
Number of employees



AI budget forecasting challenges diverge by scale

Nearly half of all organizations say the rapidly evolving AI landscape is their biggest forecasting obstacle. Enterprises struggle to keep assumptions current, while smaller firms struggle to absorb surprises. Both face real risk in a fast-moving AI market.

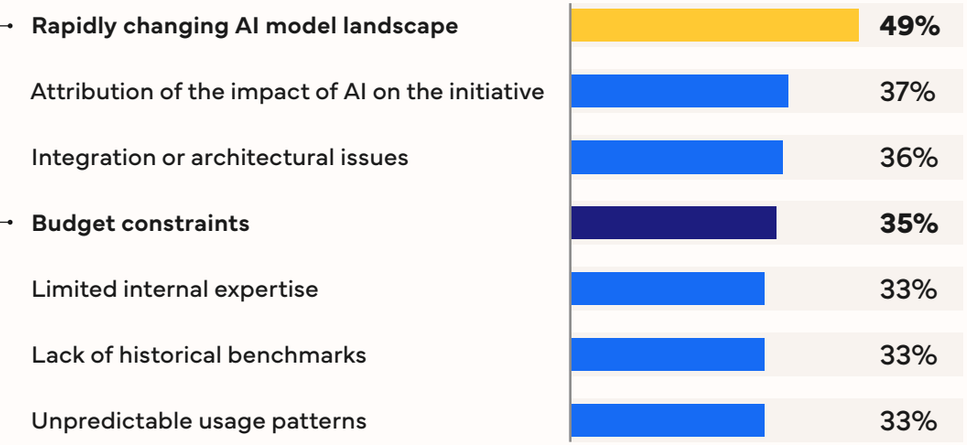


At enterprise scale, even modest updates to AI models can influence cost models across systems and regions, making adaptability and financial governance increasingly important.

As a result, **49%** of enterprises identify the rapidly evolving model landscape as a top consideration, compared with 46% of smaller businesses.
IDC defines a small business as an organization with 10–99 employees.

Smaller organizations also face a different dynamic. For **35%**, budget discipline is more challenging due to constraints, highlighting the importance of precise forecasting and right-sized investments to operate confidently within tighter margins.

What are the biggest challenges to forecasting AI initiative budgets?



n = 3,030; Source: IDC's *Microsoft Custom Survey 2026: Financial Impact of AI*

Productivity is the top priority, but ROI and cost predictability are the gates

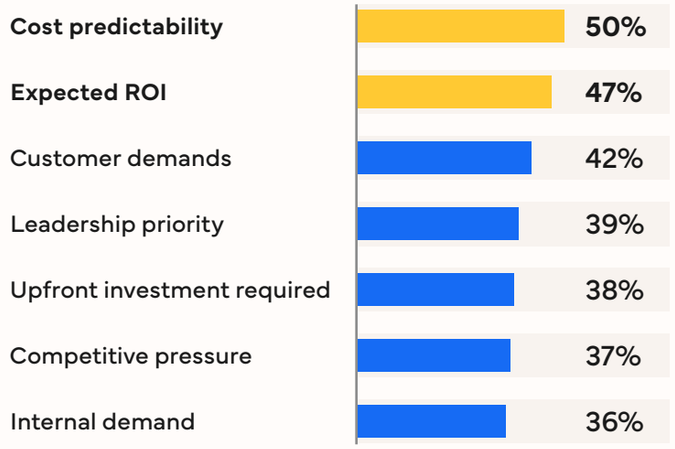
AI project sponsors that can't quantify efficiency gains will lose funding battles to those who can.

What do you feel is the primary focus of your organization's AI initiatives today?



Improving productivity as their top goal, far ahead of other AI-driven outcomes.

What are the factors that influence budget decisions for AI initiatives? (Select all that apply)



This data suggests that funding approval requires financial discipline.

1

Designing AI for efficiency:

Early choices shape long-term costs

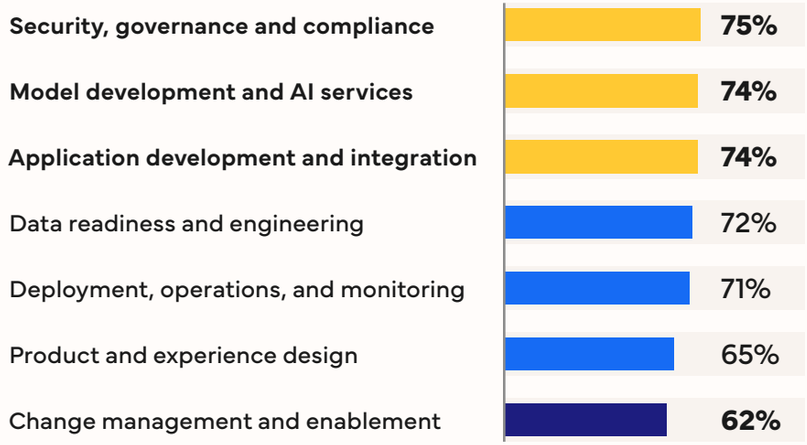
3

4

Allocation of AI budgets in the design and build stage

Security and model development receive the largest share of funding, balancing risk, scale, and speed, but change management, a critical area of AI adoption, consistently ranks last.

Budget ranking — Overall areas selected in top 5 for newly funded AI initiatives



Most of the budget flows to security, model development, and application build.

Among regulated industries:

Healthcare 78%

Automotive 77% → Prioritize protecting the investment

Manufacturing 79%

Retail 77% → Emphasize getting AI into production quickly

What stands out across every industry is what receives the least funding. Change management and enablement consistently rank last at 62%, trailing even deployment and infrastructure.

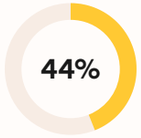


Despite heavy investment in building AI, organizations consistently underfund adoption, creating a gap between what gets built and what actually delivers ROI.

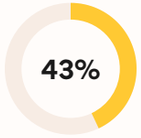
AI unit economics awareness is emerging as a core design practice

When organizations understand the cost and value of each AI interaction, they gain the ability to scale, optimize, or retire AI solutions with financial confidence.

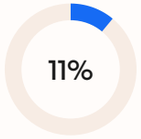
When are AI unit economics defined for a funded AI initiative?



During design



Before design starts



Defined post pilot

Most organizations now define AI unit economics (the cost and value per query, workflow, or transaction) early in the design process, with 87% doing so before or during build.

How does your organization monitor cost per outcome during design validation?



38%
Track per feature and per user segment



33%
Define outcome KPIs and cost per KPI



20%
Compare against human/manual baseline



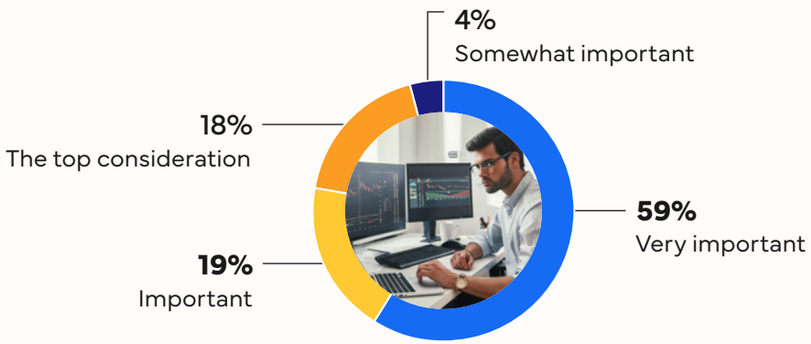
8%
Cost tracked only at project level

This awareness is increasingly supported by runtime instrumentation: About 71% of organizations track cost per outcome at the key performance indicator (KPI) or feature level rather than relying solely on total project spend, laying the foundation for active cost management as AI moves into production.

Cost-efficient models don't guarantee cost-efficient AI

There is a cost-discipline gap.

How important is cost efficiency when making the initial AI foundation model design decision?



Cost efficiency is a critical factor in AI model selection (96% cite it as important to the top consideration or very important or important), with organizations placing strong emphasis on choosing cost-effective models upfront. However, this discipline often breaks down during application design, when companies less rigorously manage long-term operational costs.

How explicitly does your organization consider long-term operating costs when designing AI applications?



Cost-capping:
We keep costs in mind and set basic guardrails.

Cost-envelope:
We design the solution to fit within an expected cost range.

Dedicated cost model:
We model how costs will evolve as usage grows.

Make-it-work:
We focus on making the solution work; cost considerations come later.

Continuous process in place:
We continuously manage and optimize costs as part of ongoing operations.

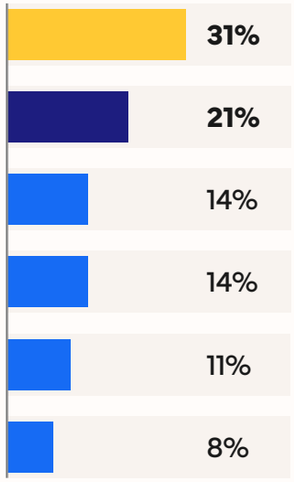
Only 17% of organizations build a dedicated cost model to project how AI costs scale with usage.

Most organizations prioritize AI application quality over cost

During AI design, organizations balance feasibility, cost, and speed to shape what they build.

These early choices directly affect scalability, operating costs, and ROI, often forcing trade-offs between long-term economics and near-term delivery and performance. Quality comes first for most. Nearly one-third of organizations design AI to hit a performance target, with cost as a secondary concern.

What best describes your organization's evaluation criteria from a designing perspective?



Quality-led: prioritize achieving target performance, with cost as a secondary constraint

Experiment-driven: use structured testing (e.g., variants or pilots) to balance cost, quality, and performance

Cost-led: prioritize minimizing cost while meeting basic quality requirements

Threshold-based: define minimum acceptable quality or performance guardrails, then optimize for cost

Risk-led: from a security, compliance, privacy, and governance perspective

Ad hoc: use evaluation criteria that vary by project with no consistent framework

Among industries:

Manufacturing
35%

Retail
36%

Quality drives design choices, where reliable performance justifies higher cost.

Energy
26%

Financial services
24%

Firms use a test-and-learn approach, favoring experiment-driven design to test performance–cost trade-offs before committing.

Smaller businesses are most exposed:
9% of SMBs rely on ad hoc design criteria, compared with 7% of enterprises, increasing the risk of cost surprises later.



1

2

Managing AI investments:

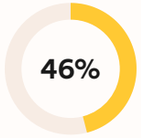
AI is controlled, but learning lags

4

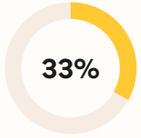
When AI hits production, accountability is centralized

Cost and performance responsibility reveal how organizations operationalize AI as a managed investment.

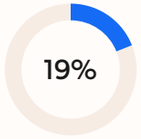
Who is primarily responsible for monitoring and optimizing AI solutions in production?



A central technology/AI platform team monitors cost and performance.



There is shared responsibility between business, technology, and finance, with defined roles and routines.



Individual teams monitor their own AI solutions informally.

Notably, 46% of organizations assign responsibility for monitoring and optimizing AI spend to a centralized IT or AI platform team.

While 33% share accountability between IT, finance, and business units.

Diffuse ownership increases the risk of not being able to control AI costs in production. Without finance on the team, organizations risk that cost data stays in operations rather than reaching the people who control the budget.

Among industries:

Telecom
39%

Energy and resources
36%

The "shared responsibility" model is most common.

Healthcare
32%

Retail
31%

The "shared responsibility" model is least common.

Healthcare
24%

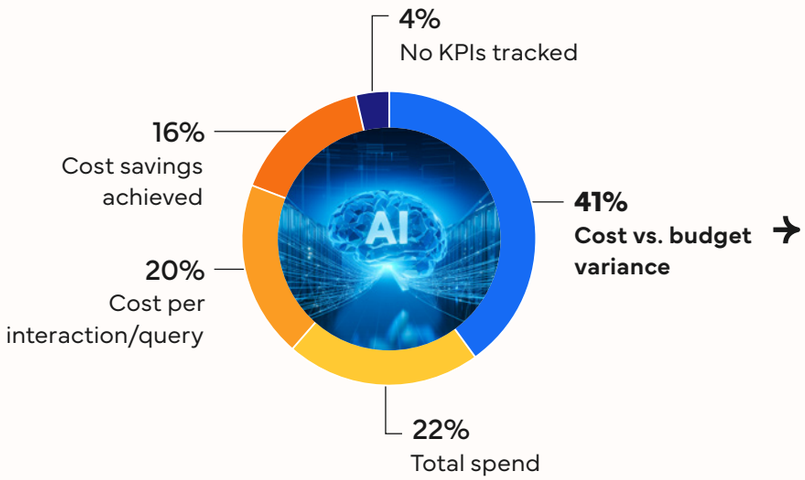
Education
24%

They compound that risk by relying on informal, decentralized monitoring at above-average rates.

Financial measures tracked for AI solutions in production

Tracking AI costs in production is common, but most metrics focus on staying within budget rather than understanding what drives value.

Which financial KPIs are actively tracked for AI solutions in production?



Among respondents, 41% rely on cost-to-budget variance as their primary KPI.

Among industries:

Budget variance is the dominant KPI across every industry.

Telecom
49%

Stands out for tracking budget variance, indicating a strong focus on staying within budget rather than on evaluating delivered value.

Financial services
23%

Telecom
23%

Energy and resources
23%

Lead on cost per interaction, reflecting familiarity with unit economics thinking.

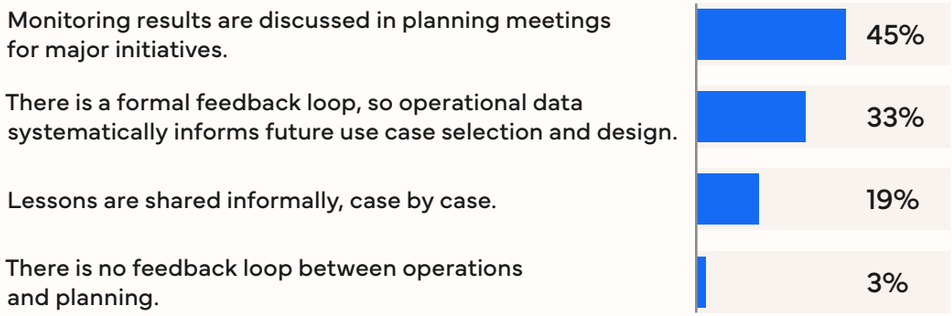
Retail
26%

Tracks total spend more than any other industry, a measure that confirms usage but offers limited insight into efficiency or value.

Production insights do not consistently shape what gets built next

Most organizations collect cost, usage, and performance data in production, but only one-third have a formal process for feeding that data back into AI planning. The rest are starting each new initiative with less information than they could have.

To what extent do monitoring insights feed back into planning for new AI initiatives?



Some regions are institutionalizing learning more rapidly through formal feedback loops.

LATAM
39%

Asia/Pacific
37%

North America
32%

EMEA
27%

Among industries:

Telecom
40%

Education
27%

Retail
27%

Enterprises are modestly ahead of small businesses, 35% versus 31%.



1 2 3

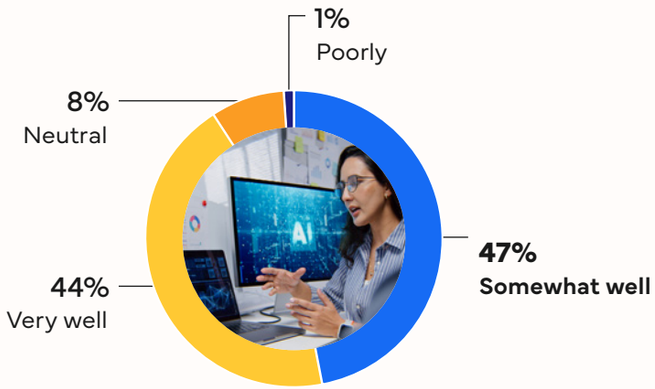
Measuring return on AI investment:

Growing focus on demonstrating value

Linking AI spend to financial impact is still maturing

With AI scaling across the business, tighter budgets are accelerating efforts to strengthen how organizations are measuring and communicating ROI.

How well is your organization able to link AI projects to clear financial outcomes?



Nearly nine in ten leaders say their organization can link AI projects to financial results, but almost half only "somewhat well."

Among industries:

Financial services
49%

Energy and resources
51%

Lead with strong financial linkage.

Healthcare
38%

Education
34%

Lag, reflecting slower and less direct paths to measurable returns.

By region:

LATAM
Linking AI to financial outcomes "very well"
52%

North America
Linking AI to financial outcomes "very well"
53%

EMEA
Skews toward "somewhat well"
53%

Asia/Pacific
Shows the highest uncertainty, "neutral"
11%

AI has become a CFO-level topic, not just a CIO project

AI investment oversight is now standard practice.

45% Strong, centralized oversight

43% Moderate oversight

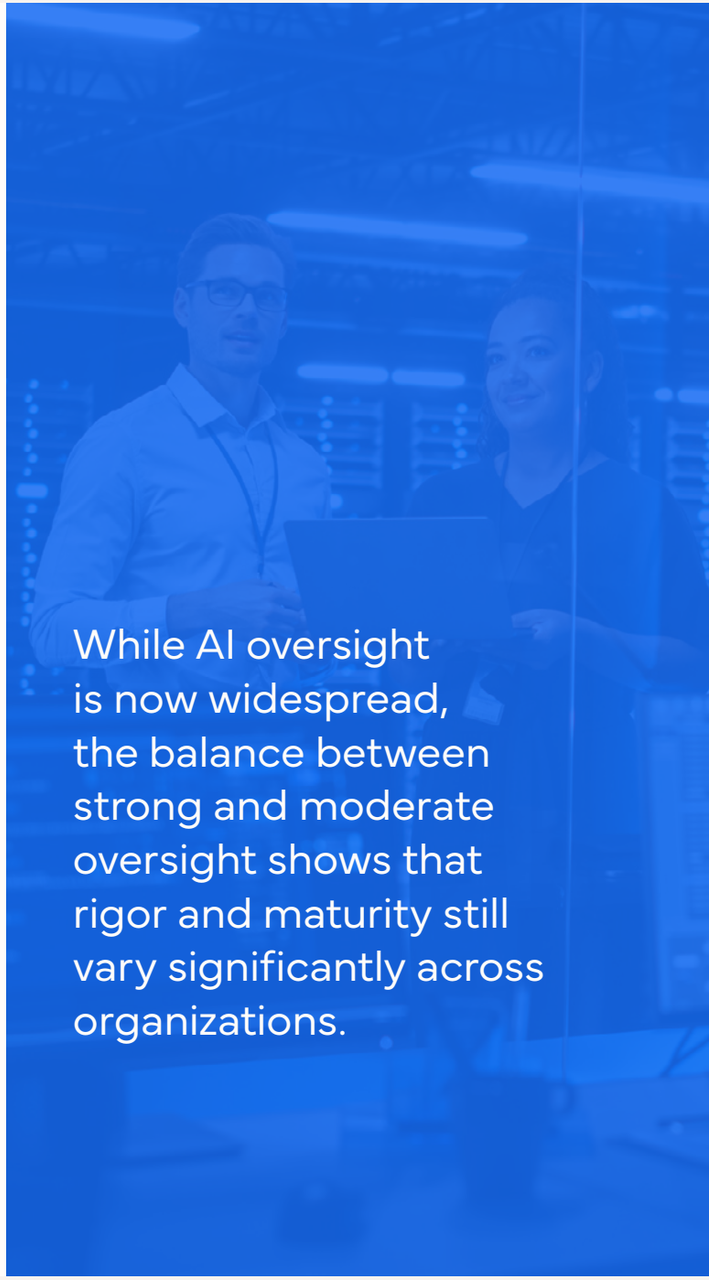
As AI costs scale in production, **nearly nine in ten organizations (88%)** report strong or moderate centralized financial oversight of AI investments.

9% Decentralized oversight

2% Minimal oversight

Only 2% operate with minimal or no oversight.

This reflects a shift from deciding whether to invest to applying the same rigor to AI investments across their life cycle as other strategic technology and capital investments.

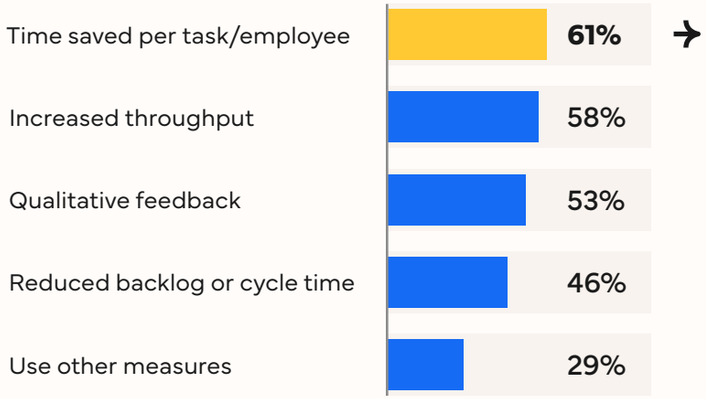


While AI oversight is now widespread, the balance between strong and moderate oversight shows that rigor and maturity still vary significantly across organizations.

Quantifying AI productivity gains distinguishes real AI value

Organizations consistently point to productivity as AI's primary source of value, and how that value is measured increasingly shapes its economic impact.

How do you quantify productivity gains from AI?



While 61% already track time saved per task, a common first step across industries, leading organizations go further by converting productivity gains into financial outcomes.

By translating time saved into overall savings, released capacity, or increased throughput, productivity shifts from a qualitative benefit to a defensible ROI. As measurement matures from perception to quantification, productivity becomes proven economic impact, not just assumed value.

Regionally, LATAM stands out in measuring throughput (75%), 17 points above the global average.

Enterprises measure more rigorously than SMBs, particularly on throughput (62% vs. 55%) and reduced backlog (48% vs. 44%), a consistent, if modest, advantage when ROI is scrutinized.

Beyond time savings, among industries:

Manufacturing
62%

→ More frequently measure output-oriented metrics (increased throughput and reduced backlog).

Financial services
65%

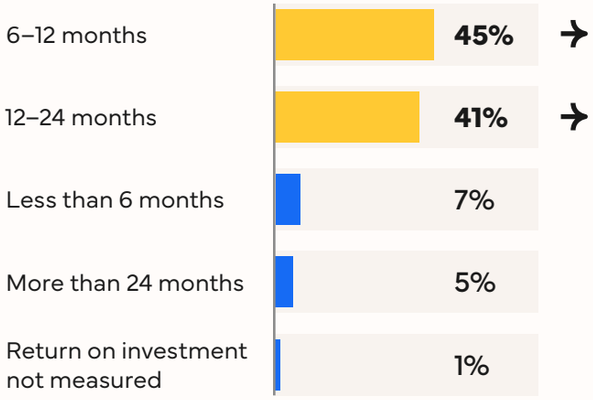
Education
65%

Most commonly measure efficiency with time saved per task.

AI operates under a defined ROI window

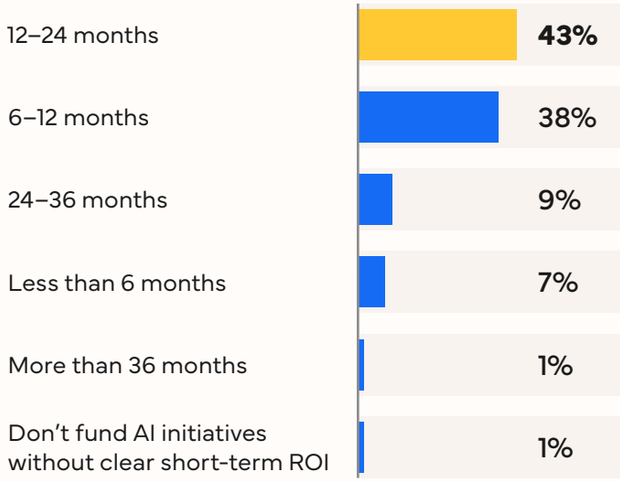
Organizations give AI initiatives a meaningful, but clearly defined, window to demonstrate financial value.

What is the expected time frame to achieve financial metric targets for AI projects?



More than 86% expect AI projects to reach financial targets within 6–24 months, signaling a strong preference for timely, measurable outcomes over prolonged evaluation cycles.

How long can your organization tolerate low ROI for an AI initiative with strong long-term potential?



This time frame also sets a practical benchmark for managing uncertainty. Even when AI initiatives offer compelling long-term strategic potential, most organizations expect ROI clarity to emerge within two years to sustain confidence and momentum.

Within this horizon, early attention to cost discipline, unit economics, and value measurement plays a decisive role. Making value visible early helps maintain funding, executive sponsorship, and the ability to scale AI initiatives successfully over the long term.



AI is delivering strategic value, with financial returns emerging

How would you rate the strategic value gained from AI investments so far?



Almost 90% of organizations rate the strategic value they gain from AI as moderate to very high.

Which types of strategic value have AI projects contributed most?

56%
Increased organizational agility

55%
Enhanced employee skills and digital maturity

53%
Improved competitive positioning

50%
Better customer trust or brand reputation

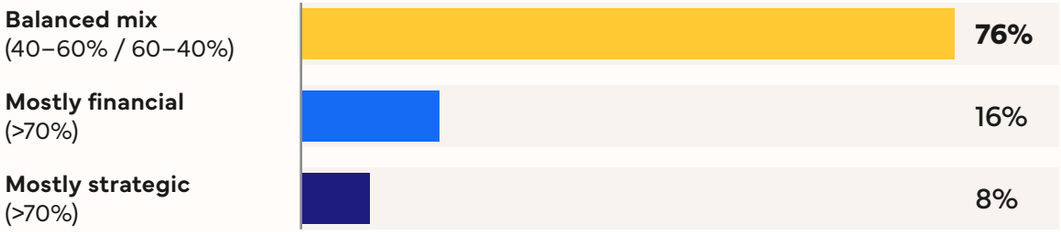
48%
Faster innovation cycles

Organizations report substantial gains in organizational agility, enhanced employee skills and digital maturity, improved competitive positioning, and faster innovation cycles. These advantages reinforce AI's role as a long-term strategic enabler and help explain why organizations continue to invest, confident that today's strategic gains will translate into sustained business and financial impact over time.

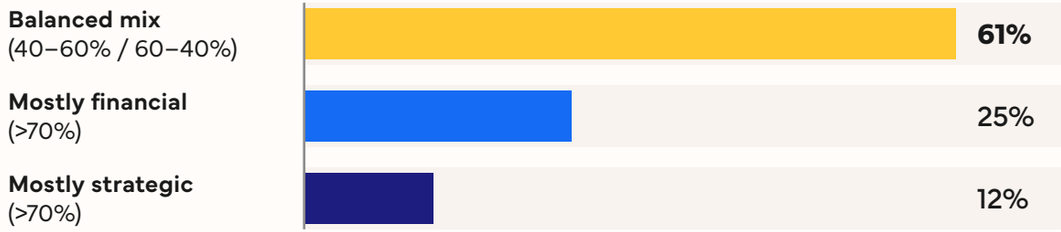
Strategic value dominates today, and financial returns are next

Organizations are making a collective bet that the strategic groundwork they lay today will translate into measurable financial returns by 2028.

What is the current share of realized AI value that is financial vs. strategic?



What is the expected share of realized AI value from financial (vs. strategic) in two years?



Today, the realized value from AI investments is overwhelmingly balanced: About 76% of organizations report a balanced mix of financial and strategic value.

But the trajectory is shifting. When asked where value will come from in two years, the share expecting mostly financial value rises significantly, from 15.7% to 25.3%, while the balanced-mix cohort shrinks from 75.7% to 61.3%.

Guidance for managing AI as an economic investment

Organizations that achieve durable AI returns build financial learning systems where every deployment sharpens cost discipline, improves measurement precision, and strengthens the next investment decision.



Extend financial discipline beyond approval.
Apply ongoing cost, usage, and performance management throughout AI's life cycle, not just at funding time.



Track AI costs at the unit level.
Add the cost per query, transaction, or workflow to governance dashboards alongside total spend to reveal how costs scale.



Design for scalable economics.
Build unit economics assumptions and operating cost models into design decisions before AI reaches production.



Feed production insights back into planning.
Use cost, usage, and performance data to inform finance and future use-case selection, not just operational tuning.



Translate productivity into financial value.
Convert saved time into released capacity or throughput gains to strengthen ROI narratives and support sustained investment under increased scrutiny.

About the IDC analyst



Kathy Lange

Research Director, AI Software, IDC

Kathy Lange is research director for IDC's AI and Automation practice, focused on machine learning life-cycle tools and technologies. Lange's core research coverage includes machine learning life-cycle technologies and platforms, trends, end-user requirements, business models, use cases, associated regulations, and market sizing for this critical, fast-growing segment. Additionally, her research covers AI life-cycle automation, model pipelines, trustworthy AI, and data annotation and labeling services.

More about Kathy Lange →

Message from the sponsor

Microsoft

Turning Bold Ambition into Real Business Advantage

As AI continues to reshape industries, the organizations pulling ahead are those that move beyond experimentation and embed AI into the core of how they operate. Realizing AI's full impact takes more than isolated pilots — it requires intelligence applied across every workflow, decision, and experience, with trust and governance built in from the start. Microsoft Frontier Transformation helps organizations make that shift, turning AI strategy into action through a trusted framework that scales adoption enterprise-wide that enriches employee experiences, reinvents customer engagement, reshapes business processes, and accelerates innovation.

Explore Frontier Transformation

IDC Custom Solutions

IDC Custom Solutions produced this publication.
The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies.

This IDC material is licensed for external use and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

IDC.com | IDC on LinkedIn | IDC Blog

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.

©2026 IDC. Reproduction is forbidden unless authorized. All rights reserved. CCPA