

[← Back to Insights](#)

STRAVORIS

AI ROI: The Measurement Crisis

Executive Summary

Enterprise AI spending is projected to reach \$2.5 trillion globally in 2026¹⁴, yet the vast majority of organizations cannot prove their AI investments deliver value. This research brief examines the widening gap between AI investment and measurable returns – a gap that has become the defining strategic challenge for technology leaders in 2026.

The data paints a consistent picture across multiple independent sources: **79% of executives perceive AI productivity gains, but only 29% can measure ROI**⁴⁷. MIT research found a **95% failure rate** for enterprise generative AI projects (defined as no measurable P&L impact within six months)². Meanwhile, only 6% of organizations report achieving AI payback within one year, with most requiring 2–4 years – far exceeding the 7–12 month expectation typical for technology investments³.

The root cause is not technological failure. It is a measurement infrastructure deficit. Organizations are attempting to evaluate AI using traditional project ROI frameworks designed for discrete software deployments. These frameworks cannot capture the compounding, cross-functional, and often intangible value that AI generates – decisions not made, errors avoided, speed gained across hundreds of workflows. The result: organizations systematically undercount AI's value, creating a feedback loop that constrains future investment precisely when scaling is most critical.

Three structural forces are converging in 2026 to make this crisis acute:

- Budget pressure:** 61% of senior leaders feel more pressure to prove AI ROI than a year ago⁸, and 53% of institutional investors expect positive returns within six months².
- The production gap:** 67% of enterprises report 101–250 proposed AI use cases, but 94% have fewer than 25 in production¹. More than two-thirds rely on manual or projected ROI tracking even for production systems¹.
- The Trough of Disillusionment:** Gartner places enterprise AI squarely in the Trough of Disillusionment throughout 2026, where "interest wanes as experiments and implementations fail to deliver"¹⁵.

The organizations breaking through – approximately 20% that Deloitte classifies as "AI ROI Leaders" – share common traits: they allocate more than 10% of technology budgets to AI, use differentiated measurement frameworks for generative vs. agentic AI, and define success in strategic terms (revenue growth opportunities, business model reimagination) rather than cost reduction alone³. The measurement crisis is solvable – but it requires abandoning the assumption that AI ROI behaves like traditional IT ROI.

Evidence Base & Methodology

Research Approach

This brief synthesizes findings from 15 primary sources gathered via targeted web searches and direct source retrieval on March 12, 2026. Searches covered seven angles: recent developments, analyst reports, counterarguments, case studies, technical perspectives, vendor landscape, and historical trajectory.

Source Profile

Source Type	Count	Examples
Industry analyst reports	4	Deloitte, Gartner, Forrester, Kyndryl
Enterprise vendor research	3	IBM, ModelOp, Cisco
Trade press / CIO publications	3	CIO.com, The Register, BizTech
Academic / independent research	3	MIT GenAI Divide, UC Berkeley, Acemoglu
Practitioner analysis	2	Ajith Prabhakar, Christian & Timbers

Date Range & Limitations

Evidence spans August 2025 to March 2026. All analyst surveys were conducted in 2025 with reports published in late 2025 or early 2026. The Deloitte survey covered 1,854 executives across 14 European and Middle Eastern countries³; geographic bias toward Europe/ME should be noted. The ModelOp survey covered 100 senior AI leaders globally¹ – a smaller but more specialized sample. MIT's 95% failure rate figure is widely cited but its exact methodology and sample size were not available from free sources. The Kyndryl survey covered 3,700 senior leaders globally⁸.

1. The Scale of the Measurement Gap

1.1 Investment vs. Returns: The Numbers

The gap between AI spending and demonstrated returns has reached a scale that demands attention at the board level. Multiple independent data sources converge on a consistent finding: most enterprises cannot prove AI value.

Metric	Statistic	Source
Global AI spending (2026)	\$2.5 trillion	Gartner ¹⁴
Cumulative U.S. AI spend (2022–2026)	>\$1.5 trillion	RBC / Bloomberg ¹⁰
ROI gap (capital deployed vs. revenue generated)	~\$600 billion	RBC / Bloomberg ¹⁰
Organizations that increased AI investment (past 12 mo.)	85%	Deloitte ³
Organizations planning to increase AI investment (next 12 mo.)	91%	Deloitte ³
Executives who perceive AI productivity gains	79%	IBM / multiple ⁴
Executives who can measure AI ROI	29%	IBM / multiple ⁴
Organizations using AI that actively measure ROI	23%	Multiple sources ¹⁰
GenAI projects with no measurable P&L impact in 6 months	95%	MIT GenAI Divide ²
AI initiatives that deliver expected ROI	~25%	IBM CEO Study ⁴
Enterprises that have scaled AI enterprise-wide	16%	IBM CEO Study ⁴

1.2 The Production Deployment Bottleneck

The measurement gap is compounded by a production gap. Organizations are generating AI use cases far faster than they can operationalize them:

- 67% of enterprises have 101–250 proposed AI use cases, yet 94% have fewer than 25 in production¹
- Only 8.6% of companies have AI agents deployed in production; 63.7% report no formalized AI initiative at all¹¹
- McKinsey finds only 23% of enterprises are scaling AI agents, while 39% remain stuck in experimentation¹¹

- Gartner warns that over 50% of enterprise AI initiatives will fail to reach production through 2027 due to missing foundational architecture¹¹

ModelOp CEO Dave Trier captured the dynamic: "Business units may hit a few singles when leadership is looking for a homerun."¹ The proliferation of use cases without production systems creates an illusion of progress that obscures the absence of measurable value.

1.3 The Timeline Mismatch

A critical source of executive frustration is the gap between expected and actual ROI timelines:

Stakeholder	Expected ROI Timeline	Source
Institutional investors	53% expect returns within 6 months	Teneo ²
CEOs	84% predict returns will take >6 months	Teneo ²
Most sophisticated management teams	Looking for returns within 12 months	IBM ²
Actual typical ROI realization	2–4 years	Deloitte ³
Organizations achieving payback within 1 year	6%	Deloitte ³
GenAI users expecting ROI within 1 year	38%	Deloitte ³

This timeline mismatch creates a dangerous dynamic: investors and boards demand quarterly proof of returns on a technology class that typically requires multi-year horizons. Organizations that cannot bridge this communication gap risk premature defunding of AI programs that are, in fact, on track.

2. Why Traditional ROI Frameworks Fail for AI

2.1 The Structural Mismatch

Among technology executives surveyed, 58% acknowledged that traditional ROI measures are insufficient for AI⁵. The reasons are structural, not cosmetic:

- **AI is not a discrete deployment.** It functions as an enterprise transformation, meaning value emerges unevenly across automation, adoption, and reinvention phases⁵.
- **Value is often intangible.** Deloitte identifies "intangible benefits" (e.g., better vendor relations, improved decision quality) as a key barrier to ROI measurement³.
- **AI is entangled with broader transformation.** Isolating AI's specific contribution from concurrent process redesign, data quality improvements, and organizational change is methodologically difficult³.
- **Technology evolves faster than metrics.** Rapidly evolving AI capabilities outpace the measurement frameworks organizations use to evaluate them³.

2.2 The Model Output vs. Business Outcome Problem

A core insight from practitioner analysis: enterprises measure **model outputs** rather than **business outcomes**⁷. Traditional AI evaluation focuses on model metrics – accuracy, precision, recall – while ignoring downstream business impact. This creates accountability gaps and prevents scaling from pilots to production.

The proposed corrective is what one analyst calls the "Decision Velocity Index" – measuring not how well the model performs, but how effectively AI-generated insights translate into organizational action⁷:

- **Decision Latency:** Time from AI output to human action
- **Override Rates:** How frequently humans reject AI recommendations
- **Time-to-Action:** Speed of workflow execution post-AI assistance
- **Decision Stream ROI:** Business impact per decision made

2.3 The Infrastructure Readiness Deficit

Measurement requires infrastructure, and most organizations lack it. The Cisco AI Readiness Index reveals significant gaps²:

Readiness Dimension	% Organizations "Fully Ready"
IT infrastructure	32%
Data preparedness	34%
Governance processes	23%

When two-thirds of organizations lack adequate data preparedness and three-quarters lack governance readiness, the measurement gap is inevitable. You cannot measure what you cannot govern, and you cannot govern what you cannot see.

3. Who Is Getting It Right – And How

3.1 The AI ROI Leaders Profile

Deloitte identifies approximately 20% of surveyed organizations as "AI ROI Leaders" – top performers who are achieving demonstrable returns³. These organizations share distinct characteristics:

Characteristic	AI ROI Leaders	General Population
Allocate >10% of tech budget to AI	95%	Not specified
Use different frameworks for GenAI vs. Agentic AI	85%	Not specified
Mandate AI training across workforce	40%	Not specified
Believe agentic AI enables strategic work	83%	Not specified
Define wins as "revenue growth opportunities"	49%	Not specified
Focus on "business model reimagination"	45%	Not specified

The critical differentiator is not spending level but **measurement sophistication**. Leaders use separate evaluation frameworks for different AI modalities and define success in strategic terms rather than pure cost reduction.

3.2 Enterprise Case Studies

Three enterprise CIOs profiled by CIO.com illustrate contrasting but effective measurement approaches²:

New York Life (Matt Marze, CIO): Evaluates AI through earnings impact – operating expense reduction, margin improvement, revenue growth, customer satisfaction, and client retention. Prioritizes AI-ready areas where data and skills already exist. This represents a *financial-first* approach anchored to P&L line items.

Palo Alto Networks (Meerah Rajavel, CIO): Selects initiatives delivering velocity, efficiency, and improved experience. Their IT operations automation improved from 12% to 75% between early 2024 and late 2025, halving operational costs. This represents an *operational velocity* approach with clear before/after metrics.

Jamf (Linh Lam, CIO): Emphasizes stakeholder goals and measurable outcome metrics. Avoids projects lacking demonstrated potential value. This represents a *stakeholder-aligned* approach that filters out

low-confidence bets early.

3.3 The Four-Domain Measurement Framework

Across sources, a consensus framework is emerging for measuring AI value across four domains, moving beyond single-metric ROI calculations⁶:

Domain	Key Metrics	Example
Operational Efficiency	Cycle time, throughput, error rate, rework %	IT automation: 12% → 75% (Palo Alto Networks)
Experience & Growth	CSAT/NPS, conversion rates, retention lifts	Customer satisfaction, client retention (New York Life)
Financial Impact	Cost-to-serve, gross margin, working capital	Operational costs halved (Palo Alto Networks)
Risk & Compliance	Policy violations avoided, audit hours saved	Cost of compliance breach offset vs. AI cost

Gartner has additionally introduced two emerging frameworks: **Return on Employee (ROE)** – measuring how AI enhances employee experience and capability – and **Return on Future (ROF)** – quantifying strategic optionality and future opportunities AI creates⁶. These attempt to capture value dimensions that traditional financial ROI inherently misses.

4. The Counterarguments and Skeptical View

4.1 The Structural Skeptics

Not everyone agrees the measurement crisis is a solvable framing problem. Nobel laureate Daron Acemoglu argues that generative AI will "at best automate profitably only 5% of all tasks," predicting a modest 0.05% annual productivity gain and a 1.1–1.6% GDP increase over 10 years – far from the doubling others forecast¹⁰. His critique centers on reliability issues, lack of human-level judgment, and inability to automate physical work.

Nvidia CEO Jensen Huang offers a different kind of caution, comparing demands for immediate AI ROI to "forcing a child to make a business plan" for a hobby – advocating instead for broad experimental exploration⁴. This view positions the measurement crisis not as a failure of measurement but as a failure of patience.

4.2 The Spending Skepticism

The aggregate numbers fuel skepticism. With cumulative U.S. AI spending projected to exceed \$1.5 trillion from 2022–2026 and an estimated \$600 billion gap between capital deployed and revenue generated¹⁰, critics question whether the industry is experiencing a bubble dynamic similar to the dot-com era. CEOs who report getting "nothing" from AI adoption efforts (56%, per one survey¹¹) lend credibility to this view.

4.3 Where Revenue Growth Remains Aspirational

Deloitte's data introduces an important nuance: while 66% of organizations report productivity and efficiency gains from AI, **revenue growth largely remains an aspiration** – 74% of organizations hope to grow revenue through AI in the future, but only 20% are doing so today³. AI investments in market cap gains also underperform: organizations investing in AI were less likely to see significant market cap gains (43%) than those investing in data (65%) or security (66%)³.

This suggests that AI's current value is concentrated in *cost reduction and operational efficiency* rather than *revenue generation* – and that the most aggressive ROI expectations (revenue growth, business model transformation) may be premature for most organizations.

5. The Governance–Measurement Nexus

5.1 Governance Platform Adoption Surge

One of the most striking data points from the ModelOp 2026 report is the surge in commercial AI governance platform adoption – from 14% in 2025 to nearly 50% in 2026¹. This tripling suggests enterprises are recognizing that governance infrastructure is a prerequisite for measurement, not an afterthought.

5.2 Agentic AI Amplifies the Challenge

The rise of agentic AI introduces new measurement complexity. Most enterprises now connect agentic AI systems to 6–20 external tools and services¹, expanding third-party risk and cost exposure. Only 10% of agentic AI users currently see significant measurable ROI, with 50% expecting returns within 3 years and 33% anticipating a 3–5 year timeline³.

Gartner predicts 40% of enterprise applications will feature task-specific AI agents by 2026, up from less than 5% in 2025¹³. This rapid adoption will create a new wave of measurement challenges: agentic systems generate value through autonomous decision chains that are harder to attribute, trace, and quantify than single-model inference.

5.3 The Self-Funding Model

TCS's Jennifer Fernandes proposes an approach that directly addresses the measurement–investment tension: use AI-driven IT modernization to fund subsequent AI initiatives through achieved efficiencies, creating self-sustaining investment cycles². This model requires rigorous measurement of early wins to justify subsequent phases – making the measurement crisis not just a reporting problem but a funding mechanism problem.

Key Assumptions & Uncertainties

1. **The 95% failure rate (MIT) is widely cited but methodologically opaque.** The definition of "failure" (no measurable P&L impact within 6 months) may be too aggressive a timeline for a technology Deloitte says typically requires 2–4 years for ROI. The true failure rate depends heavily on how "failure" is defined and over what time horizon.
2. **Geographic bias in key surveys.** Deloitte's data skews European/Middle Eastern (1,854 executives across 14 countries in those regions). U.S.-specific dynamics – including higher AI spending levels and different regulatory environments – may differ materially.
3. **Vendor survey bias.** ModelOp, IBM, Cisco, and Kyndryl all sell products that address the measurement/governance gap they document. While the directional findings are corroborated across independent sources, absolute percentages should be treated as indicative rather than definitive.
4. **The "Trough of Disillusionment" timeline is uncertain.** Gartner places enterprise AI in the Trough through 2026, but the duration and depth of any trough is inherently unpredictable. Some sectors (financial services, healthcare) may progress faster than others.
5. **Agentic AI measurement is nascent.** With only 10% of agentic AI users reporting measurable ROI³, the evidence base for measuring agentic systems is thin. Frameworks that work for generative AI may not transfer.
6. **The gap between perceived and measured value may reflect measurement immaturity, not value absence.** The 79% who perceive gains but cannot measure them may be right about the gains – the problem is the measurement tooling, not the technology. This distinction matters for investment decisions.

Strategic Implications & Actionable Insights

- 1. Invest in measurement infrastructure before scaling AI deployment.** The data consistently shows that organizations measuring AI value outperform those that do not. With governance platform adoption tripling in a single year¹, the market has recognized this – but most organizations have not yet acted. Budget for measurement tooling, data pipelines, and attribution systems as a prerequisite, not a phase-two activity.
- 2. Abandon single-metric ROI for a four-domain value framework.** Measure across operational efficiency, experience/growth, financial impact, and risk/compliance⁶. Single-metric ROI systematically undercounts AI value and creates a false narrative of underperformance. Organizations that adopted multi-dimensional frameworks are disproportionately represented among Deloitte's AI ROI Leaders³.
- 3. Shift from activity metrics to outcome metrics.** Replace "number of AI models deployed" with "cost-to-serve reduction," "decision latency improvement," and "override rate trends"⁷. The production gap (101–250 use cases proposed, <25 in production¹) is partly a consequence of measuring activity rather than outcomes.
- 4. Differentiate measurement frameworks by AI modality.** 85% of AI ROI Leaders use different evaluation approaches for generative AI vs. agentic AI³. Agentic AI's autonomous decision chains, multi-tool integrations, and longer value realization timelines demand distinct measurement approaches.
- 5. Manage the timeline expectation gap with staged milestones.** Bridge the gap between investor expectations (6-month returns²) and actual timelines (2–4 years³) by defining leading indicators – adoption rates, override rate trends, decision velocity improvements – that demonstrate trajectory before full financial returns materialize.
- 6. Use early AI wins to self-fund the next wave.** The TCS model of using AI-driven modernization savings to fund subsequent AI initiatives² addresses the measurement-investment tension directly. This requires quantifying early efficiency gains rigorously enough to justify reinvestment.
- 7. Focus AI revenue expectations on operational efficiency in the near term.** Deloitte's data shows 66% achieving efficiency gains but only 20% generating revenue growth³. Setting near-term expectations around cost reduction and operational velocity – with revenue growth as a 2–3 year horizon – aligns expectations with demonstrated outcomes.

References

1. ModelOp, "2026 AI Governance Benchmark Report Shows Explosion of Enterprise AI Use Cases as Agentic AI Adoption Surges But Value Still Lags," GlobeNewsWire, March 11, 2026. [Link](#). Accessed March 12, 2026.
2. CIO.com, "2026: The Year AI ROI Gets Real," 2026. [Link](#). Accessed March 12, 2026.
3. Deloitte, "AI ROI: The Paradox of Rising Investment and Elusive Returns," 2025–2026 (survey: Aug–Sep 2025, 1,854 executives, 14 countries). [Link](#). Accessed March 12, 2026.
4. IBM, "How to Maximize AI ROI in 2026," IBM Think, 2026. [Link](#). Accessed March 12, 2026.
5. MIT, "The GenAI Divide: State of AI in Business 2025," Summer 2025. Cited via IBM, CIO.com, and multiple secondary sources.
6. Multiple sources, "AI Value Measurement Framework – Four-Domain Model," synthesized from CIO.com, IBM, Gartner research, and practitioner analysis, 2025–2026.
7. Ajith Vallath Prabhakar, "Enterprise AI Has a Measurement Problem," March 1, 2026. [Link](#). Accessed March 12, 2026.
8. Kyndryl, "2025 Readiness Report," 2025 (3,700 senior business leaders surveyed). Cited via CIO.com and secondary sources.
9. Cisco, "AI Readiness Index," 2025–2026. Cited via CIO.com.
10. RBC / Bloomberg analysis of U.S. AI spending, 2022–2026. Cited via secondary sources including Bytelota and CIO.com.
11. Multiple sources (McKinsey State of AI, StackAI, TechRepublic), "Enterprise AI Production Deployment Statistics," 2025–2026.
12. Teneo, "CEO Outlook Survey," 2025–2026. Cited via CIO.com.
13. Gartner, "Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by 2026," August 2025. [Link](#). Accessed March 12, 2026.
14. Gartner, "Gartner Says Worldwide AI Spending Will Total \$2.5 Trillion in 2026," Gartner Newsroom, January 15, 2026. [Link](#). Accessed March 12, 2026.
15. Christian & Timbers, "Gartner AI Spending Forecast 2026 and the Renewal Era of ROI," 2026. [Link](#). Accessed March 12, 2026.

STRAVORIS

INNOVATE. INTEGRATE. ELEVATE.