

Enterprise AI ROI

What Seven Landmark Reports Found, What They Missed, and Five Decisions Worth Making Now

A Practitioner's Guide for Enterprise AI Leaders

Type: Research Synthesis | Executive White Paper

Period Covered: 2025–2026

Primary Sources: Accenture (2025) | Deloitte AI ROI Survey (Oct. 2025) | Deloitte State of AI in the Enterprise (Jan. 2026) | Google Cloud ROI of AI (2025) | McKinsey State of AI (Nov. 2025) | Microsoft Becoming a Frontier Firm (2025) | OpenAI State of Enterprise AI (2025)

Governance Framework: Factics Methodology (Fact → Tactic → KPI) | HAIA-RECCLIN | CBG v5.0

Independent Research: Puglisi, B.C., The Real State of Enterprise AI (March 2026)

How to Read This Paper

Most AI research tells you what is happening, and this paper tells you what to do about it.

Each section follows the same structure, drawn from the Factics methodology: **Fact** (what the research establishes), **Tactic** (the specific action that addresses it), and **KPI** (the measurable signal that tells you whether the action is working). The seven corporate sources provide the facts, independent research and governance methodology fill the gaps the vendors could not reach, and every section closes with a decision you can bring to your next leadership meeting.

A note on the sources. Seven of the eight sources cited here were produced by organizations with a direct commercial stake in AI adoption: Accenture, Deloitte, Google Cloud, McKinsey, Microsoft, and OpenAI. This is not a reason to discard their research, because much of it is genuinely valuable. It is, however, a reason to understand what each organization had a structural incentive not to examine. A 2025 survey captures the resulting gap precisely: 76% of organizations are deploying agentic AI systems, but only 33% maintain responsible AI controls. That 43-point spread is what happens when the research funding and the governance obligation belong to different conversations. This paper connects them.

The pattern that explains why research funding and governance obligation have historically belonged to different conversations has a name: the Economic Override Pattern, which describes how competitive pressure and shareholder return requirements predictably override safety investment when mandatory accountability structures are absent. That pattern is not a moral failing unique to any organization but an observable market dynamic that repeats across every risk domain where capability has outpaced governance, and the AI deployment record is the most recent and largest instance of it (Puglisi, 2025, Chapter 2). The five sections that follow trace the same pattern through five dimensions of enterprise AI deployment and identify the accountability structures that interrupt it.

Two modes of operating with AI. Before the five sections, one distinction is worth making explicit because it shapes how every section's tactic applies in practice.

Factory Quality means the machine checks the machine. Ethical AI establishes what the system should and should not do. Responsible AI asks who answers when the system fails. The question is intentionally unsatisfying at this tier: consequence flows back through the organization that built and deployed the system, but no individual stands at a checkpoint with personal exposure. No named human faces moral judgment from peers, employment consequence, civil liability, or criminal prosecution for what the system produced. The machine cannot answer the question, and the organization answers only in aggregate, with no human whose behavior changes because failure carries personal cost. Both Ethical AI and Responsible AI are necessary. Neither is governance.

Handmade Quality means a named human holds authority at defined checkpoints, knows where the compromises exist, and answers for what enters the world. This is AI Governance. The human faces accountability through channels no machine possesses: moral judgment from peers and profession, employment consequence for poor decisions, civil liability for negligent judgment, and criminal exposure for gross recklessness. That incentive structure is what makes governance real rather than declared.

The problem most organizations have is not that they chose Factory Quality but that they are operating under Responsible AI while representing their posture to boards and regulators as AI Governance. The five sections that follow identify where that misrepresentation creates the most measurable damage, and what closes it.

1. AI Is Capital. You Are Not Governing It Like Capital.

The Fact

Accenture's 2025 research makes the most theoretically complete economic argument in the entire corporate corpus: agentic AI is a new type of productive capital, what Accenture calls "cognitive capital," that organizations can own, manage, and compound the way traditional capital has been managed since the early days of modern business. Competitive advantages are shifting toward organizations that control agentic capital, not merely those that subscribe to AI services.

Microsoft frames the same idea operationally: the Frontier Firm "buys intelligence like electricity and compounds it like interest." Deloitte draws the structural analogy to the steam-to-electricity transition, a shift so fundamental that full benefits only emerged once organizations redesigned how they operated, not just what tools they used.

The regulatory environment is beginning to reflect this shift. FASB issued Accounting Standards Update 2025-06 in September 2025, modernizing the framework for internally developed software costs and moving disclosure requirements toward alignment with Property, Plant and Equipment rules. FASB stated explicitly that capitalization practice will not significantly change for most software under the new rules, and mandatory compliance does not begin until December 2027. The IASB has an active intangible assets project and in 2025 concluded that AI is not sufficiently different from other intangibles to warrant a separate treatment. Neither standard created a formal AI capital class, at least not yet.

What the standards have not yet required, the economics already demand. Agentic AI compounds under disciplined ownership, creates organizational dependency, and produces returns that scale with governance rigor rather than headcount. Organizations treating it as a subscription line item are making a capital allocation decision without a capital governance framework.

The gap the corporate research left open: Accenture named cognitive capital and made the case for its importance. Not one corporate report told a CFO, board, or controller what to do about it: no governance mechanism, no named accountability structure, no quarterly tracking discipline, no accounting context.

The Tactic

Before the next budget cycle, require every AI initiative above a defined materiality threshold to carry a named capital owner, meaning a specific person with documented accountability for whether that initiative compounds value over time. The classification decision itself is available now and does not require the accounting standards to catch up: require each initiative to be classified explicitly as either an operating expense or a capital investment, along with the governance obligations that follow from each classification.

Apply the CBG checkpoint architecture to each classification: *Before* the investment is approved, the named owner documents scope, expected compounding horizon, and success criteria. *During* deployment, the owner holds authority to redirect or terminate. *After* each quarter, the owner validates results against the Before criteria and reports variance to leadership with documented rationale.

This is not a technology requirement but a named-person requirement applied to decisions that currently have no named person.

The KPI

Percentage of AI initiatives above the materiality threshold carrying a named capital owner with documented quarterly accountability. Target: 100% within the first budget cycle following implementation.

Secondary signal: variance between projected and actual compounding value, reported by named owner quarterly. An owner who cannot explain variance does not have the authority the governance model requires.

2. Your ROI Is Not Reaching the Income Statement, and You Are Measuring the Wrong Thing

The Fact

Deloitte's AI ROI Survey, the only corporate report dedicated entirely to explaining why AI investment fails to reach the income statement, identifies five structural causes: benefits are intangible and hard to monetize; platforms are siloed and data quality is poor; technology evolves faster than measurement frameworks; human adoption resistance limits realized value; and AI investment is entangled with broader transformation costs that make attribution difficult.

Most organizations take two to four years to achieve satisfactory ROI on a typical AI use case, significantly longer than the seven to twelve months enterprises expect from standard technology investments. Only 6% report payback in under a year, and even among the most successful projects, just 13% see returns within twelve months.

McKinsey's contribution is the most important single data point in the corporate corpus on this question: only 39% of surveyed organizations attribute any measurable EBIT impact to AI at the enterprise level, despite those same respondents reporting genuine productivity and cost benefits at the use-case level. That gap between use-case results and enterprise P&L is not a rounding error but the place where value is structurally disappearing.

What the corporate research missed: None of the seven reports named the mechanism by which value disappears. Knowing that ROI is hard to measure does not close the measurement gap.

Four independent findings complete the picture.

The rework tax. Workday research (January 2026) found that nearly 40% of reported AI productivity gains are consumed by correcting or verifying AI-generated output. Organizations measuring gross output volume are systematically overstating AI's actual contribution to enterprise performance. Most are measuring the wrong thing without knowing it, and their metrics are confirming a story that their income statement is not telling.

Measurement misalignment. Organizations typically measure AI ROI at the use-case level, tracking individual processes and workflows, while the most revealing data, including McKinsey's, comes from the firm level. The gap between those two vantage points is precisely where value disappears: a process can run faster without the enterprise P&L moving if the time saved is reabsorbed elsewhere rather than converted into revenue growth or cost reduction. Without a named person accountable for translating use-case efficiency into firm-level outcomes, the two measurements will continue to tell different stories indefinitely.

Investment misdirection. Enterprise AI budgets concentrate in high-visibility applications: sales tools, customer-facing products, marketing automation. The highest and most consistent ROI comes from back-office automation: eliminating outsourced processes, cutting external agency spend, and streamlining operations where the cost base is controllable and the gains are directly measurable.

The shadow AI economy. A fourth structural factor operates below the governance perimeter entirely. MIT's Project NANDA found that while only forty percent of organizations hold official AI subscriptions, ninety percent of workers surveyed reported daily personal use of AI tools. That fifty-point gap represents ungoverned productivity activity invisible to enterprise metrics: drafts produced on personal devices, analyses run through consumer tools, decisions informed by AI outputs that never entered the governed workflow. Organizations measuring net productivity cannot measure what they cannot see. Shadow AI is not primarily a security problem but a measurement architecture problem, and any baseline established without accounting for it understates both the volume of AI activity and the rework that activity may be generating outside the governance perimeter. The rework tax figure of nearly forty percent may itself be conservative for organizations where shadow AI use is high.

The Tactic

Establish two measurements, not one. The first measures gross productivity at the use-case level, meaning what AI produced. The second measures net productivity, meaning what AI produced minus rework hours consumed verifying, correcting, and integrating that output. Set both baselines before the next investment cycle begins, not after.

Assign a named person at the function level to own the translation between use-case efficiency and firm-level P&L movement. This person is not responsible for the technology itself, but for connecting the efficiency gain to a line on the income statement and explaining quarterly why the connection held or did not.

Apply the Authorship Test to every productivity claim that reaches leadership: if the person presenting the productivity number cannot defend it line by line against the question "what did rework consume?", the number is not yet authored but forwarded.

The KPI

Rework hours per function per quarter, measured against the pre-investment baseline. This is the number that tells leadership whether measurement discipline is closing the gap or papering over it.

Secondary signal: percentage of use-case efficiency gains that appear as corresponding movement in firm-level P&L within two quarters. Named function owners report this number. A consistent gap between use-case results and P&L movement is a measurement architecture problem, not a technology problem, and it requires a governance response, not a technology upgrade.

3. Your Pilots Are Not Failing. They Were Never Built to Succeed.

The Fact

Deloitte's State of AI in the Enterprise (2026) describes the proof-of-concept trap precisely: a pilot runs on clean data in an isolated environment with a small team and forgiving success criteria. Nobody owns the outcome in a binding way, which is exactly why the pilot can succeed on its own terms and still produce nothing of enterprise value. Production deployment requires infrastructure investment, security reviews, compliance checks, integration with existing systems, monitoring, and ongoing maintenance, each of which demands significantly more resources and coordination than the pilot ever required.

The scale of this problem is measurable: only 25% of surveyed companies have moved 40% or more of their AI experiments into production. A healthcare AI leader quoted in the Deloitte report names the mechanism: "If there is no coherent AI strategy, you are likely to see pilot fatigue. You're chasing the next shiny object... I've seen many instances where people embark on pilots, but when asked how they'll scale up if successful, they often don't have an answer."

McKinsey confirms that nearly two-thirds of organizations have not yet begun scaling AI across the enterprise. Accenture calls the result "pilot purgatory" and identifies it as the defining failure mode of the current moment.

The number the corporate reports did not surface: S&P Global Market Intelligence found that 42% of organizations abandoned the majority of their AI initiatives before reaching production in 2025, up from 17% the prior year. The average organization scrapped 46% of its proofs-of-concept before production. That is not pilot fatigue but a structural governance failure, and the rate more than doubled in a single year.

What the corporate research missed: Every report describes the pilot-to-production gap accurately. None provides a governance gate that intervenes at the point of proposal rather than at the point of failure. Without that gate, the self-reinforcing cycle, where the next pilot always feels safer than the next scale-up, has no structural interruption.

BCG's analysis, built on work across hundreds of enterprise transformations, establishes the root cause: AI implementation success is 10% dependent on algorithms, 20% on data and technology, and 70% on people, processes, and culture. That is nearly the inverse of how most AI investment decisions are structured. Organizations spend on the technology first and discover the organizational change requirement after the spending is committed, at which point starting a new pilot is the path of least resistance.

The Tactic

Move the governance gate from the point of scale to the point of proposal. Any initiative that cannot answer four questions at proposal does not advance to pilot:

1. Who is the named production owner, meaning the specific person with binding authority over the decision to advance to production?
2. What is the integration architecture for production deployment, including the systems this must connect to?

3. What is the data quality baseline, and has it been validated against production data rather than cleansed pilot data?
4. What is the workflow redesign plan for the processes that will receive this system's output?

An initiative that cannot answer all four at proposal was never resourced to reach production, and the governance gate makes that visible before the spending is committed rather than afterward.

Apply the CBG checkpoint to the production decision specifically: the named production owner documents the advance criteria *Before* the pilot begins, holds intervention authority *During* it, and validates against those criteria *After* it completes. The board metric is not how many pilots are running but the percentage of last year's pilots now running in production.

The KPI

Percentage of prior-year pilots now running in production. Target: 40% or above within twelve months of implementing the proposal gate.

Secondary signal: average time from proposal approval to production deployment, tracked against the pre-gate baseline. If the gate is working, the initiatives that do advance should reach production faster because they entered the pipeline with their integration architecture already documented.

4. Your Employees Are Getting More Productive. Your Organization Is Not. Here Is Why.

The Fact

OpenAI's 2025 State of Enterprise AI report, drawing on data from more than 1 million business customers and a survey of 9,000 workers, finds that ChatGPT Enterprise users attribute 40–60 minutes of time saved per active day to their use of AI. Seventy-five percent of workers report completing tasks they previously could not, 87% of IT workers report faster issue resolution, and 85% of marketing and product teams report faster campaign execution.

These gains are real, credible, and consistent across multiple independent sources, and they are still not reaching enterprise P&L.

NBER Working Paper 34836, published February 2026, surveyed approximately 6,000 senior executives across the US, UK, Germany, and Australia, all at firms already using AI, and found that 90% report no measurable impact on their firm's employment or productivity after three years of AI adoption. Two-thirds of those executives personally use AI tools, averaging 1.5 hours per week, and yet nine out of ten cannot point to a firm-level result.

McKinsey identifies the correlation: AI high performers are nearly three times more likely than others to have fundamentally redesigned individual workflows, and workflow redesign has “one of the strongest contributions to achieving meaningful business impact of all the factors tested.” Microsoft confirms that making knowledge work visible, meaning understanding how work actually happens before deploying AI, is the prerequisite for redesign that holds.

What the corporate research missed: The mechanism. Why do genuine individual gains consistently fail to reach enterprise P&L even when the productivity is documented?

Two structural forces explain it. The first is private absorption: when workers gain efficiency from AI, some of that efficiency is absorbed privately, as workers use AI to improve their own work quality or reduce cognitive load rather than converting saved time into additional measured output. This is rational individual behavior, and it remains entirely invisible to enterprise metrics.

The second is the downstream bottleneck. AI-enabled workers produce more output: more code, more documents, more analyses, more drafts. The downstream processes responsible for reviewing, approving, and integrating that output were designed for pre-AI volumes. Faros AI telemetry from over 10,000 developers found that teams with high AI adoption completed 21% more tasks and generated nearly double the pull requests, while code review time for those pull requests increased 91%. The individual gain was real, and the system absorbed it because the system was not redesigned to handle the new volume.

The Tactic

Require the redesign of downstream workflows as a precondition for approving additional AI investment upstream. Do not add more AI-generated output to a pipeline that has not been redesigned to process it. The sequence matters: redesign the receiving workflow *before* expanding the upstream deployment, not after the bottleneck becomes visible as a budget problem.

Assign explicit workflow redesign authority to a named person at each function. This is a separate accountability from the production owner in Section 3 and the capital owner in Section 1. The workflow redesign owner's job is to answer one question every quarter: has the process that receives AI-generated output been redesigned to match the volume and format of what AI now produces?

Use the CBG *Before* checkpoint to document current downstream capacity before any new AI deployment is approved upstream, and use the *After* checkpoint to verify that downstream capacity increased proportionally. An organization approving upstream AI investment without a documented downstream redesign plan is producing the NBER finding by design.

The KPI

Downstream approval latency per function, measured in hours per cycle against the pre-redesign baseline. A redesign that holds produces lower latency, while a redesign that reverts produces latency returning to baseline within one or two quarters, signaling that the workflow change was not embedded.

Secondary signal: the ratio between AI output volume and downstream processing capacity, tracked quarterly by the named redesign owner. A ratio that widens is a governance signal, not a technology problem.

Beyond these operational signals, the human side of the measurement equation remains largely unaddressed in most organizations. Enterprise metrics track AI output volume, use-case efficiency, and downstream latency, but few track whether the humans governing those workflows are developing genuine collaborative capability or forwarding AI output without meaningful evaluation. One method currently being explored is the Human Enhancement Quotient, designated HEQ=AIS (Augmented Intelligence Score), which assesses the cognitive amplification produced through human-AI collaboration across four dimensions. Administered prior to AI adoption and at six-month intervals, it produces a baseline and trajectory for the human side of the productivity equation, not as a performance evaluation instrument but as a governance signal. If the collaboration produces genuine growth, the named human at the checkpoint is developing the judgment the governance architecture depends on, and if it does not, the checkpoint is forwarding rather than governing (Puglisi, 2025).

5. Two Risks Your Board Has Not Named, One With a Deadline

The Fact: Sovereign AI

Deloitte's State of AI in the Enterprise (2026) provides the most comprehensive data in the corporate corpus on sovereign AI. More than eight in ten companies, 83%, view sovereign AI as at least moderately important to their strategic planning, 77% now factor an AI solution's country of origin into vendor selection decisions, and 58% build their AI stacks primarily with local vendors. Only 21% report having a mature governance model for autonomous agents, while 74% plan to deploy agentic AI within two years.

The regulatory timeline is not theoretical. The EU AI Act's high-risk obligations for standalone AI systems go live August 2, 2026, applying to any company operating AI systems in the EU market regardless of where that company is headquartered. Penalties for prohibited AI practices reach up to 7% of worldwide annual turnover, a penalty structure that exceeds GDPR maximums. AI embedded as a safety component in regulated products such as medical devices and industrial machinery carries an extended deadline of August 2, 2027. Note on timing: as of March 2026, the EU Digital Omnibus proposal proposes to delay the Annex III deadline to December 2, 2027 and the Annex I deadline to August 2, 2028. This proposal has not been enacted, and August 2, 2026 remains the current legally binding deadline. Readers should monitor trilogue proceedings before treating the proposed dates as final.

The US CLOUD Act creates a structural legal conflict for US multinationals operating in Europe: US law permits the US government to compel US companies to produce data held anywhere in the world, which conflicts with EU data sovereignty requirements. That tension is unresolved at the treaty level as of early 2026. In practice, documented enforcement against EU-stored data by US authorities is rare, making this a legal exposure and planning risk rather than an active operational threat for most enterprises today, but it is one that must be mapped and owned.

One proposed policy approach to the CLOUD Act conflict and the broader question of AI inference accountability is the Verified AI Inference Standards Act (VAISA), submitted to the 119th Congress as part of the AI Provider Plurality Congressional Package in February 2026. The package proposes mandatory API accessibility requirements and multi-provider comparison standards as the basis for a behaviorally grounded definition of AI systems, derived from observable outputs across platforms rather than from any single developer's characterization. This is practitioner policy thinking submitted for legislative consideration rather than enacted law, and organizations should treat it as a directional signal for where US governance architecture may move rather than as compliance guidance. The underlying argument is directly relevant to organizations mapping CLOUD Act exposure today: single-provider dependence concentrates both capability and liability risk, and multi-provider architectures reduce that concentration regardless of whether any standards mandate requires it (Puglisi, 2026a).

What the corporate research missed: No corporate report distinguished between the EU AI Act's Annex III deadline (standalone high-risk AI systems, August 2026) and the Annex I deadline (AI embedded in regulated products, August 2027). Every compliance officer working in a regulated industry needs that distinction. None of the seven reports provided it, and none addressed the CLOUD Act conflict at a practical level.

The Fact: Physical AI

Fifty-eight percent of companies are already using physical AI, meaning robotics, autonomous vehicles, drones, and AI-directed physical systems, to some extent, with adoption projected to reach 80% within two years. Manufacturing, logistics, and defense lead globally, with Asia Pacific leading adoption rates.

The investment cases being built for physical AI are consistently underestimating two structural costs. The first is full deployment cost: decision-makers account for AI models and software while underestimating facility retrofits, safety infrastructure, integration with existing operational systems, maintenance contracts, spare parts, and downtime during implementation, all costs that can exceed the software investment by multiples. Physical AI deployment costs run two to three times stated estimates when facility and safety infrastructure are fully included.

The second is governance lag. Humanoid robots do not yet have a mature dedicated global safety standard for unrestricted human collaboration, while industrial collaborative robots operate under ISO 10218:2025 with collaborative operation guidance that permits work without physical separation requirements. Organizations deploying humanoid systems operate in a governance vacuum that creates genuine liability exposure, while organizations deploying industrial cobots operate under certified international safety standards. These are not the same risk profile, and business cases that treat them as equivalent are mispriced.

What the corporate research missed: Accenture, Deloitte, McKinsey, Google Cloud, Microsoft, and OpenAI collectively did not distinguish between humanoid systems and industrial collaborative robots, did not provide the ISO 10218:2025 context, and did not address the full-cost underestimation problem with enough specificity for a capital planning decision.

The Tactic

On sovereign AI: Map every AI workload against the EU AI Act's Annex III high-risk categories before August 1, 2026. The mapping requires two outputs: the category classification for each workload, and the name of the human who holds the decision authority for where that workload runs and under what sovereignty conditions. This is not a legal department task delegated to a compliance team but a named-person accountability decision that belongs on the AI governance agenda now.

Organizations that complete the mapping before July 2026 have time to remediate whatever the mapping surfaces. Organizations that do not begin now will remediate under enforcement pressure, potentially at a cost approaching 7% of worldwide annual turnover.

On physical AI: Before any physical AI commitment, humanoid or cobot alike, require a full-cost business case that includes facility retrofit costs, safety infrastructure, integration with existing operational systems, maintenance contracts, and projected downtime. Apply the CBG Before checkpoint to validate that the total is reflected in the approved budget, not just the software line. Require separate governance documentation for humanoid deployments that explicitly acknowledges the absence of a mature global safety standard and names the human accountable for the resulting liability.

The KPI

Sovereign AI: Percentage of AI workloads with a documented Annex III classification and a named decision owner for sovereignty status. Target: 100% before August 1, 2026.

Physical AI: Variance between projected and actual full deployment costs for every physical AI initiative, tracked quarterly against the pre-commitment business case. A pattern of consistent underestimation is a business case governance failure, not a project execution failure, and it requires a different remediation.

What This Means for How You Run the Next Quarter

Five decisions, each available now, each requiring no new infrastructure, and each capable of producing a measurable governance signal within one quarter of implementation.

Decision	Named Owner	Governance Action	KPI
Reclassify AI as capital	Capital owner per initiative	CBG Before / During / After on every initiative above materiality threshold	% of initiatives with named capital owner and quarterly variance report
Measure net productivity	Function-level translation owner	Net productivity baseline set before the next investment cycle	Rework hours per function per quarter vs. baseline
Gate pilots at proposal	Named production owner	Four-question proposal gate before any pilot is approved	% of prior-year pilots now in production
Redesign downstream workflows	Named redesign owner per function	Downstream capacity documented before upstream AI investment approved	Downstream approval latency vs. pre-redesign baseline
Map sovereign AI exposure	Named sovereignty decision owner	Annex III classification for every AI workload	% of workloads classified with named owner before August 1, 2026

For organizations that want a published operational starting point, one practitioner path merits attention. The five decisions above were developed through a governance methodology built and documented over several years of multi-AI workflow practice, published openly under the HAIA-RECCLIN and Checkpoint-Based Governance frameworks at basilpuglisi.com and github.com/basilpuglisi/HAIA. The methodology addresses named human authority, structured checkpoint functions, and audit trail requirements in operational terms, with version-controlled specifications and case studies available for review. It is one path, and the experience behind it is offered as evidence that the architecture works in practice, not as a claim that it is the only architecture capable of doing so. Any governance structure that assigns named humans, defines checkpoints before irreversible action, and produces accountability that survives audit across moral, employment, civil, and criminal channels satisfies the structural requirement the five decisions establish. The value of a published reference implementation is that the design decisions are visible and the failure modes are documented, which shortens the path for organizations building their own.

None of these decisions requires a new platform, a new vendor, or a new budget line, but each requires a named human with binding authority at a defined checkpoint and a measurable signal that tells leadership whether the authority is being exercised or just declared.

The organizations closing the AI value gap are not the ones that moved fastest but the ones that governed most deliberately, and the evidence for that claim is already visible in the quarterly reviews of the organizations that have not.

How to Use These Sources

Each source has a domain where it provides the strongest available evidence.

For the economic theory of AI as capital: Accenture (2025). For the governance mechanism that makes that theory actionable, including the FASB and IASB context and the CBG checkpoint architecture that converts theory into a named-person obligation: Puglisi (2026).

For diagnosing why ROI is not materializing: Deloitte AI ROI Survey (Oct. 2025), the most dedicated treatment, noting its survey was conducted across Europe and the Middle East. For the rework tax, measurement misalignment, shadow AI economy, and investment misdirection mechanisms, and the net productivity measurement framework that addresses them: Puglisi (2026), drawing on Workday, NBER, MIT NANDA, and Faros AI.

For quantifying the pilot-to-production problem: McKinsey (Nov. 2025) for enterprise-level scaling data; Deloitte State of AI (2026) for the mechanics of why pilots stall. For the proposal-stage governance gate and the 42% abandonment rate that makes it urgent: Puglisi (2026), drawing on S&P Global Market Intelligence and BCG.

For individual-to-enterprise productivity translation: OpenAI (2025) for the most detailed usage-based evidence of individual gains. For the NBER finding that those gains are not reaching enterprise P&L, and the downstream bottleneck mechanism that explains why: Puglisi (2026), drawing on NBER Working Paper 34836 and Faros AI. For measurement of the human side of the collaboration equation: Puglisi (2025), HEQ=AIS framework.

For organizational transformation methodology: Microsoft (2025) for the Frontier Firm framework and workflow redesign methodology. OpenAI (2025) for behavioral evidence from production deployments at scale.

For sovereign AI and physical AI adoption data: Deloitte State of AI (2026), the only corporate report that quantifies both. For Annex III versus Annex I regulatory distinctions, CLOUD Act enforcement reality, physical AI safety standards by system type, and proposed US policy frameworks including VAISA: Puglisi (2026).

Primary Sources

Corporate Research

- Accenture. (2025). Six Key Insights for C-Suite Executives to Maximize the Return on Agentic AI. Accenture Strategy.
- Deloitte. (2025, October). AI ROI: The Paradox of Rising Investment and Elusive Returns. Deloitte Insights. Survey: 1,854 senior executives across Europe and the Middle East.
- Deloitte AI Institute. (2026, January). State of AI in the Enterprise: The Untapped Edge. Deloitte Insights. Survey: 3,235 leaders across 24 countries.
- Google Cloud. (2025). The ROI of AI 2025. Conducted by National Research Group. Survey: 3,466 senior leaders at global enterprises.
- McKinsey & Company. (2025, November). The State of AI in 2025: Agents, Innovation, and Transformation. McKinsey Global Survey. 1,993 participants, June–July 2025.
- Microsoft. (2025). Becoming a Frontier Firm: What We Are Learning. Microsoft WorkLab.
- OpenAI. (2025). The State of Enterprise AI: 2025 Report. Based on data from more than 1 million business customers and a survey of 9,000 workers.

Independent Research

- Puglisi, B. C. (2025). The Human Enhancement Quotient: Measuring Cognitive Amplification Through AI Collaboration (HEQ=AIS Framework). basilpuglisi.com.
- Puglisi, B. C. (2025). Governing AI: When Capability Exceeds Control. ISBN 9798349677687.
- Puglisi, B. C. (2026, March). The Real State of Enterprise AI: What the Numbers Say, What Leadership Must Do. No vendor sponsorship. HAIA-RECCLIN Human-AI Governance Methodology. basilpuglisi.com.
- Puglisi, B. C. (2026a, February). AI Provider Plurality Congressional Package. Submitted to the 119th Congress. Includes Verified AI Inference Standards Act (VAISA v6). github.com/basilpuglisi/HAIA.

Additional Sources Referenced via Puglisi (2026)

- Yotzov, I., Barrero, J. M., Bloom, N., Davis, S. J., Smietanka, P., et al. (2026, February). Firm Data on AI. NBER Working Paper 34836. National Bureau of Economic Research.
- S&P Global Market Intelligence. (2025). Voice of the Enterprise: AI and Machine Learning.
- Workday, Inc. (2026, January). New Workday Research: Companies Are Leaving AI Gains on the Table.
- Faros AI. (2025). The Developer Productivity Paradox.
- BCG. (2026). From Potential to Profit: Closing the AI Impact Gap.
- MIT Project NANDA. (2025). Enterprise AI Deployment Study. Massachusetts Institute of Technology.
- FASB. (2025, September). Accounting Standards Update 2025-06. Financial Accounting Standards Board.
- IASB. (2025). Intangible Assets Project Update. International Accounting Standards Board.
- ISO 10218-1:2025 and ISO 10218-2:2025. Safety Requirements for Industrial Robots. International Organization for Standardization.
- European Commission. (2024). Regulation (EU) 2024/1689: The EU AI Act. Official Journal of the European Union.
- EY. (2025). Agentic AI Deployment and Controls Survey.
- Independent survey data on agentic AI deployment and responsible AI controls. (2025). Cited in Puglisi (2026).

Produced under the HAIA-RECCLIN Human-AI Governance Methodology and Facts framework (Fact → Tactic → KPI). Human judgment guided all synthesis decisions. All factual claims carry source attribution. #AIassisted

Basil C. Puglisi, MPA - Digital Strategy Consultant & Responsible AI² Governance
me@basilpuglisi.com | basilpuglisi.com