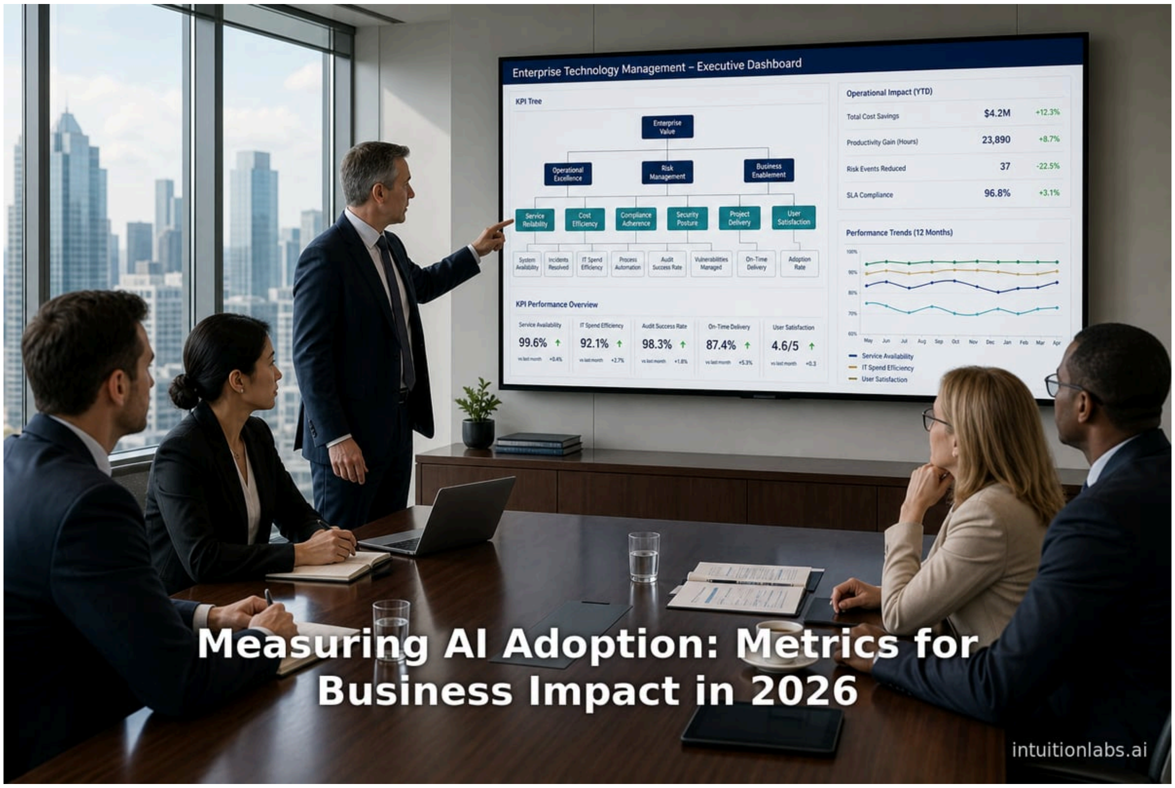
Measuring AI Adoption: Metrics for Business Impact in 2026

6/25/2026 • 45 min read

measuring ai adoption | ai adoption metrics | ai roi measurement | ai adoption framework | ai implementation kpis | ai training effectiveness | ai business impact metrics | ai competency evaluation



Executive Summary

Measuring artificial intelligence (AI) adoption has become a board-level concern precisely because most organizations still cannot do it well. As of 2026, the gap between AI usage and AI value is the defining problem in enterprise technology management. **McKinsey**'s State of AI survey (fielded June 25 to July 29, 2025, with 1,993 respondents across 105 countries) found that more than three-quarters of organizations now use AI in at least one business function (mckinsey.com), yet **PwC**'s 2026 CEO Survey, presented at the World Economic Forum's Annual Meeting in Davos, found 56% of chief executives report neither increased revenue nor decreased costs from AI investments over the prior 12 months, with only 12% reporting gains on both dimensions ([2] forbes.com) ([3] aimagazine.com). MIT's Project NANDA reported in July 2025 that despite $30 to $40 billion in enterprise generative AI (GenAI) investment, 95% of organizations were getting zero measurable return, with just 5% of integrated pilots extracting most of the value ([4] mlq.ai) ([5] fortune.com).

This report answers the central question, how does one measure AI adoption in a way that actually predicts business impact, by synthesizing survey research, vendor case studies, and government statistics into a practical measurement framework. The core finding is that adoption and impact are not the same thing, and organizations that conflate them systematically overinvest in vanity metrics such as logins and license counts. **Prosci** research found that 38% of AI adoption challenges stem from user proficiency issues, more than double the rate attributed to technical integration problems (16%) ([6] prosci.com). A workable measurement program therefore layers four distinct signal types: activation (are people using the tool), behavioral proficiency (are they using it well), business outcome (is it moving a KPI that matters), and governance (is usage safe and sustainable).

Quantitatively, adoption itself continues to climb but unevenly. **Gallup**'s Q4 2025 workplace panel (n=22,368, October 30 to November 14, 2025) found that 12% of U.S. employees now use AI daily, up from 10% the prior quarter, while frequent use (at least a few times a week) reached 26% ([7] gallup.com). The **U.S. Census Bureau**'s Business Trends and Outlook Survey found overall AI usage among firms hovering between 17% and 20% from December 2025 to May 2026, rising to 37% among firms with at least 250 employees ([8] census.gov). A Federal Reserve analysis reconciling three separate federal surveys pegged year-end 2025 firm-level adoption at about 18%, individual-level generative AI adoption at about 41%, and employment-weighted firm adoption (accounting for the fact that large employers dominate headcount) at roughly 78% ([9] federalreserve.gov).

Life-sciences companies illustrate what disciplined measurement looks like in a regulated environment. **Novo Nordisk** scaled a self-service generative AI platform to more than 25,000 employees who built over 2,500 chatbot use cases, at an average operating cost of $10 per chatbot per month ([10] aws.amazon.com). Separately, Novo Nordisk's use of **Claude Code** cut clinical study report production from more than 10 weeks to 10 minutes ([11] claude.com). Sanofi's Concierge assistant reached 72,000 monthly users by early 2026, with 90.5% of surveyed employees reporting time savings on routine tasks and a 90% positive-feedback rating ([12] aws.amazon.com). These outcome-anchored figures, tied to time, cost, and quality rather than login counts, are the model this report recommends organizations replicate. The remainder of this analysis details the adoption-versus-impact distinction, maturity-staged frameworks such as Atlassian's four-stage model, training-effectiveness measurement adapted from the Kirkpatrick model, sector benchmarks, and a practical implementation roadmap for building a defensible AI measurement program.

Introduction and Background

Artificial intelligence adoption inside organizations has outpaced organizations' ability to measure it. Generative AI tools such as ChatGPT and GitHub Copilot moved from novelty to default workplace utility within roughly three years, and by 2025 most large enterprises had at least one AI initiative in production. Yet the metrics executives use to report on that progress, active user counts, license utilization, pilot counts, mostly describe activity rather than value. This mismatch is not a minor reporting quibble. It has become the single most cited reason AI programs stall or get their budgets cut.

The scale of the problem is well documented. Gartner's November to December 2025 survey of 782 infrastructure and operations (I&O) leaders found that only 28% of AI use cases in that function fully succeed and meet return on investment (ROI) expectations, while 20% fail outright ([13] gartner.com). A separate Gartner survey of 227 chief sales officers found that 31% cited difficulty proving AI ROI as a top challenge for 2026 sales objectives ([14] gartner.com). The pattern repeats across industries: usage metrics look impressive, and financial metrics do not follow. MIT's Project NANDA, based on a systematic review of over 300 publicly disclosed AI initiatives, structured interviews with 52 organizations, and survey responses from 153 senior leaders, termed this the **GenAI Divide**: tools like ChatGPT and Copilot are widely adopted, with over 80% of surveyed organizations reporting use of general-purpose AI tools, but the vast majority of integrated pilots remain stuck with no measurable profit-and-loss impact ([15] mlq.ai) ([16] mlq.ai).

This report defines " measuring AI adoption" broadly, encompassing not just whether people log into a tool but whether that usage translates into proficiency, workflow change, and quantifiable business results. It distinguishes between **adoption metrics** (who is using AI and how often), **training effectiveness metrics** (whether people can use AI competently), and **impact or ROI metrics** (whether AI use is moving a business outcome). Treating these as one undifferentiated blob is, according to change-management researchers at **Prosci**, the primary reason organizations cannot explain what their AI investment achieved ([17] prosci.com).

The stakes for getting measurement right are compounding.As of 2026, boards increasingly demand "auditable outcomes" rather than usage anecdotes, and vendors report growing scrutiny of AI budgets ([18] forbes.com). This report surveys the current state of AI adoption data, presents the leading measurement frameworks used by consultancies and enterprises, examines named case studies with quantified outcomes, and lays out implementation guidance appropriate for organizations, including regulated industries such as pharmaceuticals and life sciences, that must balance rapid AI adoption against governance, quality, and compliance requirements.

Distinguishing Adoption Metrics From Impact Metrics

The single most consequential decision in an AI measurement program is separating "are people using this" from "is this working." Vendors overwhelmingly report the former. Legal AI platform Harvey's 2025 Year in Review disclosed more than 1,000 customers, a 92% monthly usage rate, 213.7 million files analyzed, and an 81% increase in daily-to-monthly active user ratios since launch ([19] larridin.com). These are legitimate engagement figures, but as measurement platform Larridin points out, "a 92% adoption rate tells you people are logging in, not whether they're getting more effective" ([20] larridin.com).

Prosci's research on enterprise AI adoption organizes the relevant metrics into a **four-layer measurement model** ([21] prosci.com):

- **Layer 1, Activation and access**: Are employees using the AI tool, how many active users exist, and what is license utilization. This layer serves as a baseline; a large gap between provisioned licenses and active users signals an awareness or access problem.

- **Layer 2, Behavioral change and proficiency**: Are employees working differently, and do they use AI tools effectively in real workflows. Prosci's research found user proficiency to be the single largest adoption challenge, cited by 38% of respondents, more than double the technical integration rate (16%) and organizational adoption rate (15%).
- **Layer 3, Business and operational impact**: Quality improvements, cycle time, throughput, and stakeholder satisfaction. Prosci found that team leaders' AI success indicators are dominated by performance improvements (54%), stakeholder satisfaction (13%), and time savings (12%).
- **Layer 4, Culture, governance, and trust**: Leadership clarity, transparency, and Shadow AI visibility (unauthorized tool usage). Prosci identifies leadership commitment and clarity as the single strongest differentiator between successful and unsuccessful AI adoption, with a 1.65-point gap between high and low performers, and high performers scoring +1.29 on transparency and low performers at negative 0.54.

Organizations that stop at Layer 1 risk what Prosci terms "shelfware": licensed tools that go unused in practice, where the failure is one of measurement rather than technology ([22] prosci.com). Measurement-platform vendor Larridin frames a comparable three-part model, **Utilization times Proficiency times Value**, arguing that most enterprises measure only the endpoints (spend and outcome) and skip the middle links of adoption depth and proficiency, which is exactly where the diagnostic signal lives ([23] larridin.com). Larridin also argues that power users generate 10 to 50 times more measurable value than beginners, making the proficiency gap the single largest lever available to an organization trying to improve AI ROI ([24] larridin.com). Larridin frames the underlying diagnostic question sharply: an organization that spent $4 million on AI in a given year and is asked by leadership whether it worked can typically only answer "we think so," precisely because it measured spend and outcome but not the adoption-depth and proficiency links in between ([25] larridin.com).

The practical implication is that organizations should track a small, deliberate set of metrics at each layer rather than an undifferentiated dashboard of everything measurable. A useful discipline: for every metric on a scorecard, ask which layer it belongs to and whether a leader could act differently based on a change in that number. If the answer is no, the metric is noise.

Part of the difficulty in building that discipline is structural rather than organizational. Enterprise analysis from Orange Business identifies three factors that make measuring AI impact intrinsically harder than measuring a traditional information-technology investment. First, **AI value develops over time**, since models continuously learn from data and improve their predictions, meaning the full benefits of a deployment often appear only months after go-live ([26] perspective.orange-business.com). Second, **attribution can be difficult**, because AI systems rarely operate in isolation and are typically embedded in broader workflows, making it hard to isolate how much of an observed improvement is attributable to the AI component specifically rather than to concurrent process changes ([27] perspective.orange-business.com). Third, **many benefits are intangible**: improvements in decision quality, customer satisfaction, or employee experience create real business value but are not always immediately visible in financial statements ([28] perspective.orange-business.com). These three factors, taken together, explain why organizations that rely only on quarter-over-quarter financial attribution tend to conclude AI "isn't working" even when the underlying workflow-level data, examined at the task level, shows real improvement.

The AI Adoption Maturity Framework

Because AI value does not arrive all at once, several consultancies have proposed staged maturity models that pair different metrics with different points on the adoption curve. Atlassian's Teamwork Lab, in a report published March 24, 2026, introduced the **Enterprise AI ROI Value Framework**, a four-stage model built from research the company describes as mapping "how AI value actually shows up in large organizations, from early experiments to net-new employee-facing AI tools" ([29] atlassian.com).

The four stages, and their primary metric focus, are ([30] atlassian.com):

- **Exploring**: Employees and teams experiment with AI and run pilots; the measurement focus is **adoption**. Recommended metrics include the percentage of employees experimenting with AI, tracked as daily, weekly, and monthly active users across levels and functions.
- **Optimizing**: AI is embedded into everyday workflows; the measurement focus shifts to **efficiency**, primarily time saved per task or workflow versus a pre-AI baseline.
- **Enhancing**: AI improves accuracy, consistency, and customer outcomes; the measurement focus is **quality**, including error, defect, and rework rates before versus after AI integration.
- **Transforming**: AI enables net-new products, services, and business models; the measurement focus is **innovation**, tracked through the number of AI-powered features, products, or offerings launched, and expansion into new markets or segments enabled by AI.

Atlassian researcher Ben Ostrowski summarized the transition logic behind the model: “Once you’ve nailed adoption and efficiency, the constraint isn’t how fast people can ship with AI, it’s how much you can trust what they ship. To move into the enhancing quality phase, the key is to measure AI against the KPIs you already care about, then isolate where AI-enabled changes are making a difference” ([31] atlassian.com). One chief AI officer quoted in the report overseeing more than 1,000 engineers framed the underlying pain point bluntly: “Our number-one problem right now is metrics. When it comes to AI, our board wants to understand what’s happening when the rubber meets the road” ([32] atlassian.com).

Atlassian’s own data shows that most organizations remain in the earliest stage; the report describes Exploring as the category where “most organizations are in this category” and cautions against dismissing pilot activity as “AI tourism,” since exploration is “a prerequisite to unlocking possibility” ([33] atlassian.com). This matches Gartner’s finding that AI ROI is “not driven by the sophistication of the model, but by how well the technology is integrated, governed, and aligned with real operational needs” ([34] gartner.com) and Gartner’s finding that among the 77% of I&O leaders who report at least one successful AI use case, success is attributed primarily to integrating AI into existing workflows and systems and securing full executive support ([35] gartner.com).

Failure modes are equally well documented. Of the 57% of I&O leaders in Gartner’s survey who reported at least one AI failure, the most common cause was expecting too much, too fast, with 38% citing persistent skill gaps and 38% citing poor data quality or availability as direct causes of failure ([36] gartner.com). A maturity-staged approach directly counters this by setting expectations appropriate to each phase rather than demanding Transforming-stage financial returns from a program still in the Exploring stage.

Measuring AI Training Effectiveness

If proficiency is the gating factor for AI value, as Prosci’s and Larridin’s research both indicate, then training measurement is not a peripheral human-resources function but a core part of an AI ROI program. The most widely cited adaptation is a four-level model derived from the classic **Kirkpatrick model** of training evaluation, adjusted for AI-specific behaviors.

The adapted levels, per a framework published by Pertama Partners, are:

Level	Focus	Representative Measures	Typical Timeline
1. Reaction	Learner satisfaction	Survey scores, Net Promoter Score, intent-to-apply percentage	Immediately after training
2. Learning	Knowledge and skill acquisition	Pre/post assessment score change, skill demonstration success rate	End of training

Level	Focus	Representative Measures	Typical Timeline
3. Behavior	On-the-job application	AI tool adoption rate, frequency and quality of use, manager observation	30 to 90 days after training
4. Results	Business outcomes	Productivity gain (percent), error reduction, cycle-time change, revenue or cost impact	90 to 180 days after training

Table 1 above summarizes the four-level AI training effectiveness model as adapted from Kirkpatrick's classic training evaluation framework ([37] pertamapartners.com). The critical insight in this framework is that Level 1 and Level 2 data, satisfaction scores and knowledge tests, are what most corporate training programs already collect, yet neither one predicts whether an employee will actually change how they work. The report states plainly that "completion rates and satisfaction scores aren't enough, they measure activity, not impact," and identifies behavior change as "the critical gap" where "many programs succeed at teaching skills but fail at driving application" ([38] pertamapartners.com).

Three AI-specific evaluation dimensions supplement the base Kirkpatrick structure. First, **tool adoption trajectory**: not merely whether an employee can use an AI tool but whether they consistently choose to, tracked as weekly active-usage rates across 30, 60, and 90-day windows post-training, with an expected initial spike followed by a plateau representing genuine sustained adoption. Second, **use-case expansion**: whether trained employees discover and apply AI tools to workflows beyond those explicitly covered in training, indicating deeper capability transfer. Third, **quality improvement**: whether AI-assisted work products show measurably higher quality than pre-training outputs, assessed with role-specific rubrics ([39] pertamapartners.com).

For organizations building or refining a training measurement program, the recommended sequence is: define success criteria and measurement methods before training begins; capture baseline AI skill levels, current tool usage, and current performance metrics; monitor engagement during training (completion rates, module patterns, time on content); assess knowledge acquisition immediately post-training; and track application and business results at the 30 to 90-day and 90 to 180-day marks respectively ([40] pertamapartners.com). Common failure points include skipping the baseline measurement (making improvement undemonstrable) and stopping at Level 2, since "behavior is where value emerges" ([41] pertamapartners.com).

Complementary academic work reinforces that traditional AI-literacy tests, focused on programming or statistics knowledge, often fail to predict real job performance. A November 2025 arXiv preprint on task-oriented AI literacy assessment, set in a U.S. Navy robotics training program, found a competition-style scenario task simulating real AI use outperformed conventional multiple-choice tests as a measure of applied AI literacy ([42] arxiv.org). Task-based, on-the-job assessment is thus a stronger predictor of AI competency than knowledge quizzes alone.

Academic adoption research also offers a useful corrective to purely demographic assumptions about who adopts AI fastest. An October 2025 arXiv study extending the **Unified Theory of Acceptance and Use of Technology (UTAUT)**, a widely used academic model of technology adoption, surveyed 2,257 professionals across global regions and organizational levels within a multinational consulting firm ([43] arxiv.org). The study found that organizational level significantly predicted AI adoption, with senior employees showing higher usage rates, while years of experience and geographic region were not significantly associated with adoption once organizational level was accounted for ([44] arxiv.org). This academic finding is directionally consistent with Gallup's workplace data cited elsewhere in this report, which similarly found leaders adopting AI more frequently than managers or individual contributors, and adds a research-based explanation: seniority, not tenure or location, appears to be the stronger demographic predictor of AI adoption intensity.

The scale of the underlying reskilling requirement is significant. The World Economic Forum's Future of Jobs Report 2025 found that employers expect 39% of core workforce skills to change by 2030, a figure that, while down from 44% in the 2023 edition, still represents substantial disruption, with AI and big data topping the list of technological skills projected to grow fastest in importance over the next five years ([45] weforum.org). This macro-level figure reinforces, from a workforce-planning angle, Prosci's change-management conclusion: proficiency measurement is not a one-time task but a continuous program tracking a moving target.

Evaluating AI Competency and Building an Implementation KPI Framework

Beyond training effectiveness, several organizations have proposed structured frameworks for defining and assessing individual AI competency independent of any single training event. The **CFTE AI Proficiency Framework**, published April 2026 as a public reference document, defines three public proficiency levels, six internal developmental bands, ten capability domains, and three assessment dimensions, intended to remain useful as underlying AI tools change (courses.cfte.education). The framework differentiates expectations by organizational role, noting that "leadership" needs outcome-driven goals, employee engagement, and governance oversight, while "managers" need a deep understanding of industry-specific use cases to identify risks and opportunities, and the general "workforce" needs practical training to automate routine tasks and deliver higher-quality work (courses.cfte.education). A separate European effort, the **AICET Standard**, targets standardized measurement of individual AI competencies across acculturation, advanced, and expert levels, positioning itself as more AI-specific than general digital-competence frameworks such as DigComp 2.2 (aicet.eu).

At the organizational level, Objectives and Key Results (OKRs) have emerged as a common structural device for tying AI-team output to business outcomes. The OKR Institute argues that AI transformations most often fail not on technology but on governance, citing research synthesized from multiple sources indicating that a majority of AI projects in 2025 failed to deliver intended business value, with a common root cause being a lack of clear executive alignment on success metrics, since organizations that treat AI as a pure IT initiative end up measuring "tools deployed, models trained, engineers hired" instead of outcomes ([46] okrinstitute.org) ([47] okrinstitute.org). The recommended structure cascades company-level business objectives (for example, reducing average support-ticket resolution time) into AI-team objectives with technical key results (for example, deploying automated ticket classification above a defined accuracy threshold), explicitly bridging technical metrics that do not resonate with business stakeholders and business metrics that are too distant from AI-team work to be actionable on their own (ai-solutions.wiki). An illustrative cascade in this framework pairs a company-level objective such as "deliver exceptional customer service efficiency" with business key results such as reducing average ticket-resolution time from four hours to two hours and cutting support cost per ticket from $12 to $8, alongside an AI-team objective to deploy automated ticket classification with above 92% accuracy by a defined quarter (ai-solutions.wiki).

Key implementation KPIs that recur across enterprise AI rollout guidance include:

- **Adoption depth**: Active users as a percentage of licensed seats, tracked by function and seniority level, not just in aggregate.
- **Task-level time savings**: Time to complete a defined task before versus after AI assistance, expressed in hours or minutes, not vague productivity claims.
- **Quality and error rate**: Defect, rework, or escalation rates before and after AI-assisted workflows.
- **Business-value realized**: Cash or time value attributable to AI, cross-referenced against baseline costs, described by one enterprise KPI framework as the first of five pillars alongside adoption and behavior change, model fitness, operational reliability, and risk and governance ([48] thorstenmeyerai.com).
- **Governance and safety indicators**: Compliance rate, unauthorized ("shadow") AI usage rate, and incident counts.

A recurring caution in this literature is against building sprawling, disconnected scorecards. One measurement-framework guide describes a twenty-three-metric AI portfolio dashboard as “technically correct” but “practically useless because nothing on it connects to anything else,” recommending instead a three-level KPI tree that traces from business outcome to driver to individual metric so that every number on a dashboard has a clear causal link to a result a leader can act on ([49] compelframework.org). In that framework’s terminology, a KPI tree has three levels, “the outcome level is the business result the AI programme is justified by,” which then decomposes into driver-level and metric-level measures, giving a dashboard’s numbers a causal structure rather than a flat, unranked list ([50] compelframework.org).

TechTarget’s implementation-KPI guidance, authored by Nemertes Research analysts, notes that GenAI KPIs “might require a more subjective approach” than traditional deterministic-software KPIs, since generated content must be benchmarked for creativity and relevance, not just speed and accuracy ([51] techtarget.com). The guidance offers a complementary split between **direct metrics** and **indirect metrics** that is especially useful for generative AI deployments, where subjective quality is as important as speed. Direct metrics include first-contact resolution rate (the percentage of prompts that receive a satisfactory response without modification or resubmission), content relevance score (how closely generated content matches business needs), money saved through direct and indirect cost reduction, and money made through new revenue streams the AI initiative enables ([52] techtarget.com). Indirect metrics, which matter more for generative AI than for traditional deterministic software, include customer satisfaction signals such as the frequency with which a user restates a question or expresses frustration with a chatbot’s response, user engagement rates such as session length or return frequency, and innovation scores measuring how often a system produces novel, useful outputs ([53] techtarget.com). TechTarget’s illustrative example pairs a 60% reduction in the time required to generate a personalized customer response with a parallel increase in customer satisfaction, arguing that direct and indirect metrics moving together, rather than either in isolation, is the strongest signal of a genuinely successful AI initiative ([54] techtarget.com).

Implementation Considerations and Process Changes

Standing up a credible AI measurement program requires organizational, not just technical, changes. Five considerations recur across the research reviewed for this report.

First, **executive sponsorship of the measurement function itself, not just the AI initiative**. McKinsey’s March 2025 State of AI survey found that CEO oversight of AI governance, meaning the policies, processes, and technology needed to develop and deploy AI systems responsibly, was the organizational element most correlated with higher self-reported bottom-line impact from generative AI use, particularly at larger companies, where 28% of AI-using respondents’ organizations reported the CEO as directly responsible for AI governance, particularly true “at larger companies, where CEO oversight is the element with the most impact on EBIT attributable to gen AI” (mckinsey.com) (mckinsey.com).

Second, **connecting AI usage to existing operational workflows rather than treating it as a bolt-on tool**. Gartner’s I&O survey found this to be the leading driver of success among leaders who reported at least one AI win. PwC’s 2026 CEO Survey reinforces this: CEOs reporting financial returns from AI were two to three times more likely to have embedded AI extensively across decision-making and demand generation, rather than simply distributing software licenses ([57] forbes.com).

Third, **instrumenting workflows before scaling, not after**. Forbes’ analysis of the January 2026 report cluster recommends organizations “instrument the workflow” by using admin dashboards to map consumption to specific teams and correlate that consumption with output KPIs, and to “fund the redesign” of processes rather than only funding software licenses, since ROI correlates most strongly with how embedded AI becomes in a workflow ([58] forbes.com).

Fourth, **defining success differently at each organizational level**. Prosci's research found that executives skew toward strategic and operational outcomes such as cost savings and new growth opportunities, team leaders balance tactical delivery with strategic requirements, and frontline employees are motivated by practical task-level improvements such as faster response times ([59] prosci.com). A measurement program that reports only one of these three perspectives will look successful to one audience and hollow to the other two, functioning like a balanced scorecard that evaluates AI success across organizational, team, and individual dimensions rather than relying on a single adoption metric.

Fifth, **anticipating a widening gap between technology leaders and laggards**. Gallup's Q4 2025 data show total AI use in the technology industry reaching 77%, including 57% frequent users and 31% daily users, compared with 33% total use in retail, including 19% frequent users and 10% daily users ([60] gallup.com). Gallup also found that employees in remote-capable roles reached 66% total AI use by 2025, including 40% frequent and 19% daily users, versus 32% total use among non-remote-capable roles. Organizations should expect that a single blended adoption number will obscure large internal disparities that matter for where training and governance investment should go.

Finally, life-sciences and other regulated organizations face an additional layer of complexity: measurement programs must also capture compliance and validation status, not solely productivity, since an unmeasured or unvalidated AI workflow in a regulated process carries regulatory as well as financial risk. IntuitionLabs' review of pharma and biotech AI adoption found that while roughly half of pharma and biotech companies were leveraging AI or big data in research and development by 2025 to 2026, and nearly 85% of large pharmaceutical companies considered AI an "immediate priority," only 13% to 30% of companies reported a comprehensive AI program fully implemented, with the majority still testing or planning ([61] intuitionlabs.ai). That gap between stated priority and full implementation is itself a measurement finding: intent-to-adopt metrics, if reported uncritically as adoption, systematically overstate real organizational readiness. The capital backing this adoption push is substantial and itself independently measurable: Silicon Valley Bank's 2026 Healthcare Investments and Exits report found that AI-focused healthcare companies attracted more than $18 billion in investment in 2025, representing 46% of all healthcare investment for the year, even as total healthcare investment fell 12% year over year to $46.8 billion ([62] svb.com). That combination, AI's investment share rising while overall sector investment contracts, is itself a useful proxy metric: capital allocators are treating AI-enabled healthcare and life-sciences companies as a distinct, better-differentiated category worth measuring separately from the sector at large.

Data Analysis and Evidence

The quantitative record on AI adoption is large, drawn from overlapping but methodologically distinct surveys, and shows meaningfully different numbers depending on unit of analysis (individual versus firm), weighting (headcount versus firm count), and question wording (AI broadly versus generative AI specifically). Reconciling these differences is itself part of responsible measurement practice.

A consistent finding across independent research firms is that AI adoption and AI maturity are not the same measurement, and the gap between them is large. Accenture's 2025 survey of 2,000 companies found that 70% recognize the importance of proprietary data to scaling AI, yet only 15% of companies were assessed as "AI reinvention-ready," having built the essential capabilities needed to scale AI successfully, and only 8% qualified as true "front-runners" who are scaling AI effectively and embedding it into core business strategy ([63] accenture.com). The performance gap between these groups is measurable in financial terms: front-runners with annual revenue exceeding $10 billion grew revenue 7 percentage points faster than companies still experimenting with AI, and delivered shareholder returns 6 percentage points higher across company sizes ([64]

accenture.com). Even among this elite group, front-runners had scaled only 34% of their identified strategic AI bets, indicating even the most advanced organizations view their own AI scaling programs as unfinished ([65] accenture.com). Accenture also found that companies scaling just one strategic AI bet were nearly three times more likely to exceed their ROI expectations, and that C-suite sponsorship made success 2.4 times more likely, both figures reinforcing the executive-oversight finding from McKinsey's research cited earlier in this report ([66] accenture.com) ([67] accenture.com).

Function-specific adoption data shows a similarly rapid pace of change within a single year. Salesforce's early findings from its State of Service research, based on 3,075 service professionals surveyed across 13 countries, found that AI agent adoption in customer service jumped from 39% to 66% in a single year ([68] salesforce.com). A jump of that magnitude in a single function underscores why organization-wide, once-a-year adoption benchmarks quickly go stale, and why the measurement cadence recommended throughout this report, tracking adoption quarterly or more frequently rather than annually, matters in practice and not just in theory.

At the firm level, the U.S. Census Bureau's Business Trends and Outlook Survey (BTOS), reviewing the six months from December 14, 2025 to May 3, 2026, found overall AI usage hovering between 17% and 20%, with 20% to 23% of businesses expecting to adopt AI within the following six months ([8] census.gov). Adoption correlated strongly with firm size: 37% of firms with at least 250 employees reported using AI, versus 32% of firms with 100 to 249 employees, and less than 20% of firms with four or fewer employees ([69] census.gov). By sector, the Information sector (42% expected adoption) and Finance and Insurance (39% expected) led, while Retail Trade lagged at 14% current and 17% expected usage ([70] census.gov). AI use also barely changed among the smallest firms during this period even as larger firms kept climbing ([71] census.gov). International data from the Organisation for Economic Co-operation and Development (OECD) shows the same size gradient outside the United States: in 2025, 52% of large firms across OECD member countries reported using AI, compared with only 17.4% of small firms (digital-skills-jobs.europa.eu).

A Federal Reserve analysis published April 3, 2026 reconciled the BTOS with two other federal data sources: the individual-level Real-Time Population Survey (RPS) and the Atlanta Fed's Survey of Business Uncertainty (SBU). The reconciliation found that firm-level BTOS adoption stood at about 18% at year-end 2025; work-related generative AI adoption reported by individuals in the RPS reached about 41% as of November 2025; and the employment-weighted SBU estimate, which accounts for large employers dominating headcount, found that 78% of the labor force works at a firm that has adopted AI in some form, and 54% at a firm using large language models specifically ([9] federalreserve.gov). The note attributed this wide variance primarily to differences in sampling distributions and units of analysis, alongside possible effects of question framing, information asymmetries between respondent types, and social desirability bias ([72] federalreserve.gov).

Table 2 below compares individual-level workplace AI usage across the primary current sources, illustrating why a single "adoption rate" headline figure is inherently ambiguous without specifying source, population, and definition.

Source	Population	Metric	Value	As of
Gallup Panel	U.S. employees	Daily AI use at work	12%	Q4 2025
Gallup Panel	U.S. employees	Frequent use (few times/week or more)	26%	Q4 2025
Gallup Panel	U.S. employees	Any use at work (few times/year or more)	51% (49% report never using AI)	Q4 2025
Census BTOS	U.S. firms (all sizes)	Currently using AI in any business function	17 to 20%	Dec 2025 to May 2026
Census BTOS	U.S. firms, 250+ employees	Currently using AI	37%	May 2026

Source	Population	Metric	Value	As of
Federal Reserve (RPS)	U.S. individuals	Work-related generative AI adoption	~41%	Nov 2025
Federal Reserve (SBU)	U.S. labor force (employment-weighted)	Works at a firm that has adopted AI	~78%	Nov 2025

The Gallup Panel figures come from a biweekly Gallup Panel survey of 22,368 respondents fielded October 30 to November 14, 2025, with a margin of error of plus or minus 1.0 percentage points at the 95% confidence level. Nearly half of U.S. workers, 49%, told Gallup they "never" use AI in their role as of Q4 2025 ([7] gallup.com), a figure worth stating alongside the more commonly cited adoption percentages because it underscores that a majority of the U.S. workforce, in this particular survey design, remains outside even occasional AI use.

Organizational-level adoption, distinct from individual tool usage, also shows a leveling pattern. Gallup found that in Q4 2025, 38% of employees said their organization had integrated AI technology to improve productivity, efficiency, and quality, essentially unchanged from the prior quarter, while 41% said their organization had not implemented AI tools and 21% did not know ([73] gallup.com). Leadership-level adoption consistently outpaces the rest of the workforce: 69% of leaders reported using AI at least a few times a year in Q4 2025, compared with 55% of managers and 40% of individual contributors, and frequent use among leaders rose from 17% in Q2 2023 to 44% in Q4 2025, versus a rise from 9% to 23% among individual contributors over the same period ([74] gallup.com).

On the financial-outcome side of the ledger, the picture is more sobering than the usage figures alone suggest. PwC's 2026 CEO Survey of business leaders found that 56% report neither revenue increase nor cost decrease from AI investments in the prior 12 months, 30% report increased revenue, 26% report decreased costs, and only 12% report both simultaneously ([75] forbes.com) ([3] aimagazine.com). MIT's Project NANDA, drawing on interviews across 52 organizations and 153 senior-leader survey responses collected at four major industry conferences between January and June 2025, found 95% of organizations getting zero measurable return on generative AI investment against $30 to $40 billion in enterprise spend, with just 5% of integrated pilots extracting most of the realized value ([76] mlq.ai) ([4] mlq.ai). Deloitte's 2026 State of AI in the Enterprise report found worker access to AI rose 50% during 2025, and that the number of companies with 40% or more of AI projects in production was expected to double within six months, with 53% of surveyed organizations citing enhanced insights and decision-making, 40% citing reduced costs, and 38% citing enhanced client relationships as the primary business functions AI serves ([77] deloitte.com) ([78] deloitte.com). Only 20% of the same surveyed organizations cited improving products and services and fostering innovation as a primary function AI serves, notably lower than the operational and cost-oriented use cases, suggesting most enterprise AI deployments as of 2026 remain concentrated in Atlassian's Optimizing stage of the maturity framework described earlier in this report rather than the Transforming stage ([79] deloitte.com). Deloitte frames the underlying tension for leaders bluntly as a shift from "productivity vs. reimagination" ([80] deloitte.com).

Software engineering, one of the earliest AI-assisted functions to generate controlled experimental data, offers a useful benchmark for what rigorous productivity measurement looks like. GitHub's controlled experiment comparing developers using GitHub Copilot against a control group found that Copilot users completed a defined coding task 55% faster on average (1 hour 11 minutes versus 2 hours 41 minutes), with a statistically significant result (P=.0017) and a 95% confidence interval for the speed gain of 21% to 89%. The Copilot group also had a higher task-completion rate, 78% versus 70% for the control group (github.blog). The peer-reviewed paper underlying this experiment, published on arXiv by Peng, Kalliamvakou, Cihon, and Demirer, reports that recruited software developers asked to implement an HTTP server in JavaScript as quickly as possible completed the task 55.8% faster with Copilot access than the control group, with the authors noting that observed heterogeneous effects show particular promise for helping less experienced developers transition into software careers ([81] arxiv.org). This example illustrates the kind of task-level, controlled measurement, comparing a defined task's completion time and success rate with and without AI assistance, that the broader literature recommends organizations replicate internally rather than relying solely on self-reported survey data.

Practitioner sentiment outside the consultancy research above is notably more skeptical about how widespread rigorous measurement actually is. On the finance forum r/FPandA, one practitioner wrote, "From what I've seen, very few enterprises are genuinely measuring AI ROI with any rigour. Most teams track adoption or productivity anecdotes rather than hard financial outcomes," adding that measurement works best "when companies tie AI initiatives to clear operational metrics like cost per transaction, time to insight, or labour hours saved" ([82] reddit.com). A function-specific survey confirms this at scale: SHRM surveyed 1,908 HR professionals in December 2025 and found 56% do not formally measure AI investment success at all, and only 16% use a defined ROI metric ([83] shrm.org).

Case Studies and Real-World Examples

Novo Nordisk: Self-Service Generative AI at Enterprise Scale

Novo Nordisk, a Denmark-headquartered pharmaceutical company employing more than 64,000 people across 80 countries ([84] mongodb.com), began experimenting with large language models and generative AI in 2023 and built a self-service platform on Amazon Bedrock so employees could build and customize chatbots without needing to develop or maintain their own infrastructure. Senior vice president Jens Jepsen explained the rationale: "A lot of colleagues had great ideas for how to solve their own business problems using generative AI, but they were limited by the fact that they couldn't develop the applications or maintain the infrastructure to support them" ([85] aws.amazon.com).

By the time of AWS's published case study, more than 25,000 Novo Nordisk employees had used the platform to create chatbots for over 2,500 unique use cases, spanning document drafting, information retrieval, and virtual-colleague or critic functions ([10] aws.amazon.com). More than 1,000 employees used the company's general-purpose off-the-shelf chatbot, which processed over 26,000 prompts per month and was trained on 140,000 documents for its largest use case, moving from proof of concept to minimum viable product in one month ([86] aws.amazon.com). Cost measurement was deliberately kept simple: each use case runs on serverless AWS infrastructure at an average cost of $10 per month ([87] aws.amazon.com).

Separately, Novo Nordisk deployed Anthropic's Claude Code for clinical documentation and drug-development workflows, reducing the time spent producing clinical study report documentation from more than 10 weeks to 10 minutes, and cutting resources needed for device verification protocols by 95% ([11] claude.com) ([88] claude.com). These figures, sourced from a vendor case study, should be read as vendor-reported outcomes rather than independently audited results, but they are directionally consistent with time-savings figures reported elsewhere in Novo Nordisk's AI program and illustrate the kind of task-specific, before-and-after time measurement that this report's earlier sections recommend as a best practice.

A third, independently documented Novo Nordisk deployment reinforces the same pattern from yet another vendor relationship. Working with Microsoft's AI Acceleration Studio, Novo Nordisk's FounData initiative, which harmonizes more than 200,000 patient-years of clinical trial data aligned to open industry standards including CDISC, SDTM, and ADaM, built a governed reasoning agent on Azure that lets researchers analyze clinical data while maintaining rigor, oversight, and compliance ([89] microsoft.com). Sid Prabhu, who leads the FounData initiative, said "the pharmaceutical industry is under enormous pressure to improve productivity" as research spend rises while competition accelerates ([90] microsoft.com), and described the measurable before-and-after shift in team capacity directly: "In the past, we might have had the capacity to pursue 5 to 10 strong ideas in a quarter. Now we can evaluate 50-plus" ([91] microsoft.com). That five-to-tenfold increase in evaluated ideas per quarter is a concrete capacity metric, distinct from the adoption and time-saved metrics discussed elsewhere, and illustrates a fourth measurable category: how many more decisions an organization can afford to evaluate at the same staffing level.

Sanofi: From Productivity Tool to Digital Gateway

Sanofi built an internal AI companion called Concierge on Amazon Bedrock, initially as part of its Digital Accelerator initiative. As adoption crossed 20,000 users, Sanofi's leadership reclassified Concierge from a productivity tool into the company's primary digital gateway for accessing AI-powered agents and internal systems, replacing the need to navigate thousands of legacy applications ([92] aws.amazon.com).

As of the first quarter of 2026, Concierge reached 50,000 weekly users and 72,000 monthly users, generating 11 million conversations with a 90% positive-feedback rating. Among surveyed employees, 90.5% reported saving time on routine tasks, 78% said they made better decisions, and 77% said Concierge improved their work-life balance ([12] aws.amazon.com). In one specific use case cited by Sanofi, an employee used Concierge to cut verification-task time by 92% ([93] aws.amazon.com). Sanofi head of people services Matthieu Lachieze framed the connection between adoption and impact directly: "Concierge is a critical tool that empowers Sanofi to transform itself faster. We're not only becoming more efficient but also improving the quality of what we deliver for patients and what our employees experience" ([94] aws.amazon.com).

Pfizer: Search-Time Reduction Through the PACT Initiative

Pfizer, which serves an estimated 1.3 billion patients globally, partnered with AWS on the Pfizer-Amazon Collaboration Team (PACT) initiative, pursuing 14 projects combining generative AI and machine learning to accelerate digital drug development. Pfizer's published results include saving scientists up to 16,000 hours of search time annually and cutting infrastructure costs by 55% ([95] aws.amazon.com). This case illustrates a different measurement pattern than the Novo Nordisk and Sanofi examples: rather than reporting broad adoption figures, Pfizer's disclosed metrics are narrowly scoped to two specific, auditable categories, aggregate search-time saved and infrastructure cost, which are more directly attributable to specific AI tooling decisions.

AstraZeneca: Multi-Agent Decision Support in Clinical Development

AstraZeneca, pursuing a stated goal of delivering 20 new medicines by 2030, deployed a multi-agent AI tool called Development Assistant, built on Amazon Bedrock Agents, that lets clinical, regulatory, safety, and quality teams ask natural-language questions and receive actionable insights from structured and unstructured data in seconds rather than through manual analysis ([96] aws.amazon.com). While AstraZeneca's public case study emphasizes qualitative acceleration of decision-making rather than a single headline percentage, it is a useful counterpoint to the other cases in this section: not every credible AI deployment reports a discrete ROI figure, and organizations should be cautious of dismissing a program as unmeasured simply because it has not yet produced a single summary statistic; the underlying task-level time and quality metrics may still exist even if not yet publicly aggregated.

(Hypothetical Example) A Mid-Size Life-Sciences Commercial Team Applying the Four-Layer Model

To illustrate how the Prosci four-layer measurement model might apply outside the largest global pharmaceutical companies, consider a hypothetical mid-size specialty-pharma commercial organization piloting an AI-assisted field-insights tool. In the Exploring stage, the team would track the percentage of field representatives who log into the tool weekly (Layer 1, activation). Within 60 to 90 days, it would shift attention to whether representatives who use the tool draft call summaries independently and with fewer manager corrections (Layer 2, proficiency). At the Optimizing and Enhancing stages, the team would measure whether average call-summary turnaround time drops and whether summary error rates, verified against source CRM records, decline (Layer 3, business impact). Throughout, the team would track whether usage stays within approved, compliant workflows rather than shifting to unsanctioned consumer AI tools (Layer 4, governance). This hypothetical sequence mirrors the staged logic of Atlassian's Enterprise AI ROI Value Framework and Prosci's four-layer model described earlier in this report, and is presented here only as an illustrative composite, not as a description of any specific organization.

Implications and Future Directions

Several trends visible in the current data are likely to shape how organizations measure AI adoption over the next 12 to 24 months. First, the gap between usage metrics and financial metrics that MIT's NANDA report and PwC's CEO Survey both documented is unlikely to close through better AI models alone; Gartner's finding that ROI depends on integration and governance rather than model sophistication suggests the bottleneck is organizational, not technical ([34] gartner.com). Boards and CFOs are likely to keep raising the evidentiary bar for AI budget renewal, meaning measurement programs built today should anticipate audit-level scrutiny rather than anecdotal reporting.

Second, the proficiency gap identified by Prosci and quantified by Larridin, where power users may generate 10 to 50 times more value than beginners, implies that organizations will increasingly need role-specific, not organization-wide, proficiency targets and coaching programs ([24] larridin.com). A single company-wide "AI adoption rate" is likely to become less useful to executives over time as a management tool, replaced by function-level and even individual-level proficiency dashboards.

Third, as agentic AI systems (AI that can take multi-step autonomous action rather than respond to single prompts) become more common, measurement frameworks will need new categories beyond the largely conversational-assistant metrics reviewed in this report. Deloitte's 2026 enterprise AI report already flags that "AI agents outpace their guardrails" as a distinct trend area, separate from productivity measurement ([97] deloitte.com), suggesting that governance and safety metrics, not just productivity metrics, will take on greater weight in adoption scorecards. IBM's own AI adoption research cites Gartner's prediction that by 2028, 33% of enterprise software will include agentic AI, up from under 1% in 2024, and notes AI-specific governance roles grew 17% in 2025, a sign organizations are already staffing for agentic oversight ([98] ibm.com) ([99] ibm.com). IBM's guidance further cautions that "some companies start ambitious AI programs without identifying clear operational goals," the same core diagnosis Gartner, Prosci, and MIT reach independently ([100] ibm.com).

Broader macro-level data corroborates the usage acceleration underneath these measurement challenges. Stanford HAI's 2025 AI Index Report found 78% of organizations reported using AI in 2024, up from 55% the prior year, alongside $33.9 billion in global generative AI private investment, an 18.7% increase from 2023 ([101] hai.stanford.edu). Read against the more conservative Census and Federal Reserve figures cited earlier, this gap itself demonstrates the report's central argument: Stanford's figure reflects any reported AI usage internationally, Census reflects a narrower U.S. definition, and neither number is "wrong," they answer different questions.

Fourth, regulated industries such as pharmaceuticals face a distinctive measurement challenge: adoption and impact metrics must be reconciled with validation and audit-trail requirements that general enterprise frameworks do not address. IntuitionLabs, an advisory firm working with life-sciences organizations on Veeva ecosystem and AI implementation, and a Veeva Vault CRM X-Pages partner, notes in its own service description that AI adoption in this sector must be paired with "strategic guidance on digital transformation, AI adoption, and technology roadmapping" that accounts for regulatory compliance from the outset rather than retrofitting it after a tool is already in production use ([102] intuitionlabs.ai). This perspective is consistent with the broader finding that governance metrics (Prosci's Layer 4) are not an optional add-on for regulated sectors but a precondition for any adoption or impact figure to be trustworthy in the first place.

Finally, measurement methodology itself is likely to keep converging toward the pattern demonstrated by GitHub's controlled Copilot experiment: task-level, before-and-after, statistically evaluated comparisons rather than aggregate self-reported survey data. As more organizations build internal capability to run this kind of controlled measurement on their own workflows, the reliability of AI ROI claims across the industry should improve, narrowing the current gap between the small number of organizations reporting rigorous, auditable results and the larger number reporting only usage anecdotes.

Vendor-side usage telemetry is itself becoming more granular in ways that should help this convergence. Anthropic's January 2026 Economic Index report introduced "economic primitives," foundational measures spanning user and AI skill level, task complexity, autonomy granted to the AI, and task success rate, derived from anonymized analysis of Claude conversations in November 2025 ([103] anthropic.com). The report found that the ten most common work tasks represented 24% of observed consumer usage, up from 21% in January 2025, while for first-party enterprise API customers, task concentration was even higher, with the top ten tasks representing 32% of traffic ([104] anthropic.com). This kind of task-level usage concentration data, published directly by a frontier AI vendor rather than inferred from survey self-report, is the type of granular telemetry that Larridin's utilization-proficiency-value framework and Prosci's four-layer model both argue organizations need in order to move past login counts toward genuine behavioral and task-level measurement. The same report also found that augmentation, conversations where a user learns, iterates, or gets feedback from the AI system, again exceeded fully automated use on Claude.ai, with augmented conversations jumping five percentage points to 52% of the sample and automated use falling four points to 45% ([105] anthropic.com), a distinction that itself functions as a measurement category: whether AI use is best understood as a collaborative aid or a replacement for a task matters for how an organization should account for the resulting productivity change.

Frequently Asked Questions (FAQs)

What is the difference between AI adoption and AI impact? AI adoption describes whether and how often people use an AI tool, typically measured through login counts, active-user rates, or license utilization. AI impact describes whether that usage produces a measurable business outcome, such as time saved, error reduction, or revenue change. Larridin defines these explicitly: "AI Adoption: The rate at which employees and teams begin using artificial intelligence tools and technologies," while "AI Impact: The measurable business value and productivity improvements that result from AI implementation" ([106] larridin.com).

How do you measure AI training effectiveness? The most widely used framework adapts Kirkpatrick's four levels, reaction, learning, behavior, and results, to AI-specific behaviors such as sustained tool-adoption trajectory, use-case expansion beyond the original training content, and measurable quality improvement in AI-assisted work products, assessed at defined intervals from immediately post-training through 90 to 180 days later.

How do you measure the ROI of AI? Rigorous ROI measurement requires connecting a defined AI spend to a defined, auditable business outcome rather than relying on adoption counts alone, using KPIs such as first-contact resolution rate, content relevance, money saved, and money made ([52] techtarget.com). Analysts also recommend instrumenting workflows to map consumption data to specific teams and correlating that with output KPIs, and funding process redesign rather than only software licenses, since embedded, workflow-integrated AI use correlates far more strongly with financial return than access alone.

What is a good AI adoption framework for a large organization? Staged maturity frameworks, such as Atlassian's four-stage Exploring, Optimizing, Enhancing, and Transforming model, or Prosci's four-layer Activation, Behavior, Impact, and Governance model, are preferable to a single flat scorecard because they match the metric being tracked to the organization's actual maturity stage, avoiding the common failure IBM identifies, where "some companies start ambitious AI programs without identifying clear operational goals," leaving them unable to define meaningful success metrics for a program still in early pilot phases ([100] ibm.com).

How do you evaluate AI competency in employees? Task-based, scenario-driven assessment tends to outperform conventional knowledge tests as a predictor of real-world AI competency, as demonstrated in a 2025 arXiv study of a U.S. Navy robotics training program, where a competition-style scenario task outperformed adopted or self-developed multiple-choice tests ([42] arxiv.org). Standardized frameworks such as CFTE's AI Proficiency Framework and the AICET Standard offer structured, role-based competency ladders as an alternative or complement to organization-built assessments (courses.cfte.education).

What are the most important AI implementation KPIs? Recurring KPI categories across the frameworks reviewed in this report are adoption depth (active users as a percent of licensed seats, by function), task-level time savings, quality and error rate, business-value realized in cash or time terms, and governance indicators such as compliance rate and unauthorized-tool usage rate. One enterprise KPI white paper frames the underlying diagnosis bluntly: "Most AI programs fail not because the models are weak, but because leaders track the wrong things (or nothing at all)," proposing a five-pillar operating view covering business value, adoption and behavior change, model fitness, operational reliability, and risk and governance that a team can review weekly ([107] thorstenmeyerai.com) ([48] thorstenmeyerai.com).

Does higher AI adoption automatically mean higher AI proficiency? No. Individual-level usage data can rise even as an organization's average skill with the tool stays flat, which is why frameworks reviewed in this report consistently separate a usage or activation metric from a distinct proficiency or behavior-change metric. The OKR Institute's guidance on AI transformation makes the same point at the organizational level, noting that pure technology metrics such as "tools deployed, models trained, engineers hired" fail to capture whether an organization is winning with AI, versus outcome metrics such as "decisions improved, revenue generated" ([108] okrinstitute.org).

Conclusion

Measuring AI adoption in 2026 requires more discipline than counting logins or licenses. The evidence surveyed in this report, spanning McKinsey, MIT, Gartner, PwC, Deloitte, Gallup, the U.S. Census Bureau, the Federal Reserve, Prosci, Atlassian, and multiple named enterprise case studies, converges on a consistent conclusion: usage is rising steadily but unevenly, while the financial return organizations can demonstrably attribute to that usage remains elusive for the majority. The organizations that close this gap share a common pattern. They separate activation metrics from proficiency metrics from business-outcome metrics rather than blending them into a single vague "AI success" narrative. They stage their expectations to match their actual maturity level, whether that is Atlassian's Exploring stage or Transforming stage. They measure training effectiveness through sustained behavior change rather than completion rates. And they build governance and compliance measurement into the program from the start rather than retrofitting it, a consideration that carries particular weight in regulated sectors such as pharmaceuticals and life sciences.

None of the frameworks reviewed here promise a single number that captures AI's full value. Instead, they offer a disciplined way of connecting the metrics an organization already has the ability to collect, adoption counts, proficiency assessments, task-level time and quality data, to the outcomes that ultimately justify continued investment. As AI tooling continues to evolve toward more autonomous, agentic systems, the specific metrics tracked will keep changing, but the underlying principle established across this body of research is likely to persist: adoption is a leading indicator, not a result, and organizations that measure only the leading indicator will keep mistaking activity for value.

External Sources

[1] https://www.mckinsey.com/\~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/2025/the-state-of-ai-how-organizations-are-rewiring-to-capture-value_final.pdf?ref=filipeoliveira.pt#:~:More%...

[2] https://www.forbes.com/sites/guneyyildiz/2026/01/28/56-of-ceos-see-zero-roi-from-ai-heres-what-the-12-who-profit-do-differently/#:~:The%2...

[3] https://www.aimagazine.com/news/what-pwc-ceo-survey-reveals-about-roi-on-ai-investments#:~:The%2...

[4] https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf#:~:Despi...

[5] https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/#:~:But%2...

[6] https://www.prosci.com/blog/measuring-ai-success-in-organization-wide-adoption#:~:Prosc...

[7] https://www.gallup.com/workplace/701195/frequent-workplace-continued-rise.aspx#:~:U.S.%...

[8] https://www.census.gov/library/stories/2026/05/ai-use-businesses.html#:~:The%2...

[9] https://www.federalreserve.gov/econres/notes/feds-notes/monitoring-ai-adoption-in-the-u-s-economy-20260403.html#:~:This%...

[10] https://aws.amazon.com/solutions/case-studies/novo-nordisk-generative-ai-case-study/#:~:Using...

[11] https://claude.com/customers/novo-nordisk#:~:10%2B...

[12] https://aws.amazon.com/solutions/case-studies/sanofi-bedrock-case-study/#:~:As%20...

[13] https://www.gartner.com/en/newsroom/press-releases/2026-04-07-gartner-says-artificial-intelligence-projects-in-infrastructure-and-operations-stall-ahead-of-meaningful-roi-returns#:~:Only%...

[14] https://www.gartner.com/en/newsroom/press-releases/2026-05-19-gartner-survey-shows-thirty-one-percent-of-chief-sales-officers-cited-difficulty-proving-roi-of-ai-driven-tools-as-a-top-challenge-for-sales-objectives-in-two-thousand-twenty-si#:~:A%20G...

[15] https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf#:~:Just%...

[16] https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf#:~:Tools...

[17] https://www.prosci.com/blog/measuring-ai-success-in-organization-wide-adoption#:~:Organ...

[18] https://www.forbes.com/sites/guneyyildiz/2026/01/28/56-of-ceos-see-zero-roi-from-ai-heres-what-the-12-who-profit-do-differently/#:~:The%2...

[19] https://larridin.com/blog/ai-adoption-vs-impact-metric#:~:Accor...

[20] https://larridin.com/blog/ai-adoption-vs-impact-metric#:~:But%2...

[21] https://www.prosci.com/blog/measuring-ai-success-in-organization-wide-adoption#:~:The%2...

[22] https://www.prosci.com/blog/measuring-ai-success-in-organization-wide-adoption#:~:Organ...

[23] https://larridin.com/workflow-mapping/how-to-measure-ai-roi-beyond-surveys?hs_amp=true#:~:Real%...

[24] https://larridin.com/workflow-mapping/how-to-measure-ai-roi-beyond-surveys?hs_amp=true#:~:The%2...

[25] https://larridin.com/workflow-mapping/how-to-measure-ai-roi-beyond-surveys?hs_amp=true#:~:Your%...

[26] https://perspective.orange-business.com/en/measuring-ai-impact-how-to-turn-ai-initiatives-into-measurable-business-value/#:~:First...

[27] https://perspective.orange-business.com/en/measuring-ai-impact-how-to-turn-ai-initiatives-into-measurable-business-value/#:~:Secon...

[28] https://perspective.orange-business.com/en/measuring-ai-impact-how-to-turn-ai-initiatives-into-measurable-business-value/#:~:Third...

[29] https://www.atlassian.com/blog/ai-at-work/enterprise-ai-roi-framework#:~:We%20...

[30] https://www.atlassian.com/blog/ai-at-work/enterprise-ai-roi-framework#:~:Atlas...

[31] https://www.atlassian.com/blog/ai-at-work/enterprise-ai-roi-framework#:~:%E2%8...

[32] https://www.atlassian.com/blog/ai-at-work/enterprise-ai-roi-framework#:~:%E2%8...

[33] https://www.atlassian.com/blog/ai-at-work/enterprise-ai-roi-framework#:~:Most%...

[34] https://www.gartner.com/en/newsroom/press-releases/2026-04-07-gartner-says-artificial-intelligence-projects-in-infrastructure-and-operations-stall-ahead-of-meaningful-roi-returns#:~:A%3A%...

[35] https://www.gartner.com/en/newsroom/press-releases/2026-04-07-gartner-says-artificial-intelligence-projects-in-infrastructure-and-operations-stall-ahead-of-meaningful-roi-returns#:~:To%20...

[36] https://www.gartner.com/en/newsroom/press-releases/2026-04-07-gartner-says-artificial-intelligence-projects-in-infrastructure-and-operations-stall-ahead-of-meaningful-roi-returns#:~:Facto...

[37] https://www.pertamapartners.com/insights/measuring-ai-training-effectiveness-metrics-that-matter#:~:Train...

[38] https://www.pertamapartners.com/insights/measuring-ai-training-effectiveness-metrics-that-matter#:~:Compl...

[39] https://www.pertamapartners.com/insights/measuring-ai-training-effectiveness-metrics-that-matter#:~:Three...

[40] https://www.pertamapartners.com/insights/measuring-ai-training-effectiveness-metrics-that-matter#:~:Befor...

[41] https://www.pertamapartners.com/insights/measuring-ai-training-effectiveness-metrics-that-matter#:~:Skipp...

[42] https://arxiv.org/pdf/2511.05475#:~:We%20...

[43] https://arxiv.org/abs/2510.15142v1#:~:we%20...

[44] https://arxiv.org/abs/2510.15142v1#:~:Organ...

[45] https://www.weforum.org/stories/2025/01/future-of-jobs-report-2025-jobs-of-the-future-and-the-skills-you-need-to-get-them/#:~:Emplo...

[46] https://okrinstitute.org/ai-transformation-okrs/#:~:Most%...

[47] https://okrinstitute.org/ai-transformation-okrs/#:~:When%...

[48] https://thorstenmeyerai.com/insights/real-kpis-from-enterprise-ai-rollouts/#:~:What%...

[49] https://www.compelframework.org/articles/building-a-kpi-tree-for-an-ai-program#:~:The%2...

[50] https://www.compelframework.org/articles/building-a-kpi-tree-for-an-ai-program#:~:The%2...

[51] https://www.techtarget.com/searchenterpriseai/tip/How-businesses-can-measure-AI-success-with-KPIs#:~:A%20g...

[52] https://www.techtarget.com/searchenterpriseai/tip/How-businesses-can-measure-AI-success-with-KPIs#:~:More%...

[53] https://www.techtarget.com/searchenterpriseai/tip/How-businesses-can-measure-AI-success-with-KPIs#:~:Indir...

[54] https://www.techtarget.com/searchenterpriseai/tip/How-businesses-can-measure-AI-success-with-KPIs#:~:Consi...

[55] https://www.mckinsey.com/\~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/2025/the-state-of-ai-how-organizations-are-rewiring-to-capture-value_final.pdf?ref=filipeoliveira.pt#:~:Our%2...

[56] https://www.mckinsey.com/\~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/2025/the-state-of-ai-how-organizations-are-rewiring-to-capture-value_final.pdf?ref=filipeoliveira.pt#:~:That%...

[57] https://www.forbes.com/sites/guneyyildiz/2026/01/28/56-of-ceos-see-zero-roi-from-ai-heres-what-the-12-who-profit-do-differently/#:~:This%...

[58] https://www.forbes.com/sites/guneyyildiz/2026/01/28/56-of-ceos-see-zero-roi-from-ai-heres-what-the-12-who-profit-do-differently/#:~:%2A%2...

[59] https://www.prosci.com/blog/measuring-ai-success-in-organization-wide-adoption#:~:Organ...

[60] https://www.gallup.com/workplace/701195/frequent-workplace-continued-rise.aspx#:~:Line%...

[61] https://intuitionlabs.ai/articles/ai-adoption-pharma-biotech-benchmarks#:~:Howev...

[62] https://www.svb.com/news/company-news/ai-investment-accounted-for-nearly-half-of-healthcare-investment-in-2025-silicon-valley-bank-releases-17th-healthcare-investments-and-exits-report/#:~:With%...

[63] https://www.accenture.com/us-en/insights/ai-data/front-runners-guide-scaling-ai#:~:Only%...

[64] https://www.accenture.com/us-en/insights/ai-data/front-runners-guide-scaling-ai#:~:Front...

[65] https://www.accenture.com/us-en/insights/ai-data/front-runners-guide-scaling-ai#:~:But%2...

[66] https://www.accenture.com/us-en/insights/ai-data/front-runners-guide-scaling-ai#:~:Compa...

[67] https://www.accenture.com/us-en/insights/ai-data/front-runners-guide-scaling-ai#:~:AI%20...

[68] https://www.salesforce.com/plus/experience/agentforce_world_tour_nyc_spring_2026/series/best_of_agentforce_world_tour_nyc_spring_2026/episode/episode-s1e31#:~:AI%20...

[69] https://www.census.gov/library/stories/2026/05/ai-use-businesses.html#:~:For%2...

[70] https://www.census.gov/library/stories/2026/05/ai-use-businesses.html#:~:The%2...

[71] https://www.census.gov/library/stories/2026/05/ai-use-businesses.html#:~:Betwe...

[72] https://www.federalreserve.gov/econres/notes/feds-notes/monitoring-ai-adoption-in-the-u-s-economy-20260403.html#:~:The%2...

[73] https://www.gallup.com/workplace/701195/frequent-workplace-continued-rise.aspx#:~:Organ...

[74] https://www.gallup.com/workplace/701195/frequent-workplace-continued-rise.aspx#:~:Leade...

[75] https://www.forbes.com/sites/guneyyildiz/2026/01/28/56-of-ceos-see-zero-roi-from-ai-heres-what-the-12-who-profit-do-differently/#:~:The%2...

[76] https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf#:~:Metho...

[77] https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html#:~:Worke...

[78] https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html#:~:Enhan...

[79] https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html#:~:Impro...

[80] https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html#:~:2%2E%...

[81] https://arxiv.org/abs/2302.06590#:~:Abstr...

[82] https://www.reddit.com/r/FPandA/comments/1ox5w0r/formally_measuring_gen_ai_roi_skeptical/#:~:Hey%2...

[83] https://www.shrm.org/topics-tools/research/state-of-ai-hr-2026/full-report#:~:The%2...

[84] https://www.mongodb.com/solutions/customer-case-studies/novo-nordisk#:~:Novo%...

[85] https://aws.amazon.com/solutions/case-studies/novo-nordisk-generative-ai-case-study/#:~:Novo%...

[86] https://aws.amazon.com/solutions/case-studies/novo-nordisk-generative-ai-case-study/#:~:Simil...

[87] https://aws.amazon.com/solutions/case-studies/novo-nordisk-generative-ai-case-study/#:~:Each%...

[88] https://claude.com/customers/novo-nordisk#:~:95%25...

[89] https://www.microsoft.com/en/customers/story/26569-novo-nordisk-as-azure#:~:The%2...

[90] https://www.microsoft.com/en/customers/story/26569-novo-nordisk-as-azure#:~:%22Th...

[91] https://www.microsoft.com/en/customers/story/26569-novo-nordisk-as-azure#:~:%E2%8...

[92] https://aws.amazon.com/solutions/case-studies/sanofi-bedrock-case-study/#:~:Emplo...

[93] https://aws.amazon.com/solutions/case-studies/sanofi-bedrock-case-study/#:~:Beyon...

[94] https://aws.amazon.com/solutions/case-studies/sanofi-bedrock-case-study/#:~:%E2%8...

[95] https://aws.amazon.com/solutions/case-studies/pfizer-PACT-case-study/#:~:Under...

[96] https://aws.amazon.com/solutions/case-studies/astrazeneca-case-study/#:~:As%20...

[97] https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html#:~:5%2E%...

[98] https://www.ibm.com/think/insights/ai-adoption-challenges#:~:Gartn...

[99] https://www.ibm.com/think/insights/ai-adoption-challenges#:~:AI-sp...

[100] https://www.ibm.com/think/insights/ai-adoption-challenges#:~:Some%...

[101] https://hai.stanford.edu/ai-index/2025-ai-index-report#:~:In%20...

[102] https://intuitionlabs.ai/#:~:Advis...

[103] https://www.anthropic.com/research/anthropic-economic-index-january-2026-report#:~:This%...

[104] https://www.anthropic.com/research/anthropic-economic-index-january-2026-report#:~:The%2...

[105] https://www.anthropic.com/research/anthropic-economic-index-january-2026-report#:~:In%20...

[106] https://larridin.com/blog/ai-adoption-vs-impact-metric#:~:%2A%2...

[107] https://thorstenmeyerai.com/insights/real-kpis-from-enterprise-ai-rollouts/#:~:Most%...

[108] https://okrinstitute.org/ai-transformation-okrs/#:~:AI%20...

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at https://intuitionlabs.ai/contact for a consultation.

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by Adrien Laurent, a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.