
MEASURING THE ENTERPRISE ROI FROM AI

How time, workflow, quality, and outcome data reveal value from AI-assisted work.

Amanda Finch

Director, Risk and Compliance

Journyx, Inc.

September 2026

The Question Behind the Investment

Organizations are under pressure to understand their return on investment (ROI) in AI. But that return is often obscured. It cannot be established merely by counting licenses, measuring adoption, or asking employees whether AI saves them time. Organizations must determine whether AI changes the labor required for specific work; whether output, speed, quality, or capacity improves; and whether those operational changes produce measurable business value. If work takes less time, what happens to the hours released?

The first step toward understanding AI ROI is not calculating a sweeping company-wide number. It is selecting a consequential unit of work, establishing how much time and effort it requires today, and measuring what changes when AI enters the workflow.

Time data are the connective tissue between AI-assisted work and enterprise value. It connects AI use to changes in labor, workflow performance, and enterprise value. Time data must be analyzed alongside other data to prove AI ROI. But when the expected return depends on changes in labor effort, workflow capacity, project economics, or cycle time, credible time data are essential to establishing what changed and whether it created value.

The AI ROI Problem

Organizations are investing heavily in AI tools. Executives expect productivity gains, and employees often report that AI saves time. But Finance often cannot reliably map those claims to economic value. Leaders are left relying on adoption statistics, anecdotes, and vendor estimates. Some process-mature enterprises have performed the analysis needed to measure AI returns and redesign workflows around the technology. Many other organizations still struggle to know where to start, particularly when time and money are limited.^{1,2}

Adoption data establish that AI is being used; they do not establish that its use is creating economic value.

Usage statistics can show whether employees are using AI. Alone, they cannot show value. Measuring value requires evidence that the organization is shifting to more valuable work, lowering costs, improving quality, increasing useful output, or serving customers better.

Organizations are using AI at several degrees of operational integration. In AI-assisted work, a person initiates the task, directs the AI, evaluates its output, and remains responsible for the result. In AI-enabled workflows, AI performs one or more defined workflow steps while people retain control over

material decisions and handoffs. In agentic workflows, AI may plan and execute multiple steps, interact with other systems, and take actions with greater autonomy. These categories do not represent a mandatory sequence. Although organizational AI adoption is now widespread, AI-agent deployment remains early across most business functions.³

This paper is for organizations beginning to evaluate returns from AI-assisted work or AI-enabled workflows. As a general operating principle, an organization should establish its ability to measure and govern AI-assisted work before giving AI consequential operational autonomy. If it cannot define the unit of work, establish a baseline, measure changes in labor effort, and connect those changes to actual value while people remain in control, it will be poorly positioned to evaluate or govern a more autonomous system.⁴

Why There Is No Single ROI Metric

There is no shortcut to surfacing an organization's actual return from AI adoption. A practical approach begins with data that many organizations already gather, including time data. For AI investments that materially affect labor or workflow performance, time data connect AI use to changes in effort, output, cost, and business outcomes.

Conventional Approaches and Their Limitations

Some common approaches to measuring AI ROI have limitations.

License-cost comparison. *"AI costs \$X per employee, and employees say it saves Y hours."* This comparison seems logical at a glance. But self-reported time savings are usually estimates of varying accuracy. While employees may spend the time saved on other work, it may be lower-value work. Also, the benefits of time saved vary widely by role and workflow; this simple comparison can't account for that variation.

AI adoption and usage metrics. Although it makes sense to ask whether AI is actually being used, usage alone is not enough. Organizations that track AI logins, prompts, active users, and generated content are measuring utilization, not results. What is the point of adopting AI if it fails to generate value? Usage tracking might be useful in the early days after AI introduction, but understanding ROI demands further investigation.

Output measurement. Increased output is promising, especially when production time does not rise. But more code, documents, reports, or customer responses do not necessarily create a commensurate increase in value or quality.

Headcount reduction. Treating labor elimination as the only valid, or even primary, return overlooks possible gains in capacity, responsiveness, innovation, and revenue. It may also encourage employees to conceal AI-enabled efficiencies.

Employee surveys. Surveys are another flavor of self-reporting. They may be useful for identifying workflows that can best benefit from AI. But they should only generate hypotheses rather than serve as final ROI evidence.

A Layered Approach to Seeing AI ROI

Measuring AI ROI requires several layers of evidence. No single measure establishes the return; organizations must connect the layers relevant to their operations. The framework below applies established principles of productivity measurement and AI evaluation to workflow-level adoption.⁵

Measurement layer	Question answered	Examples
Investment	What did AI cost?	Licenses, implementation, training, governance
Usage	Where is AI being used?	Users, tools, workflows, frequency
Time	Did the amount of labor required change?	Hours per task, project, client, or process
Output	Did capacity or throughput change?	Cases handled, reports produced, releases completed over a period of time
Quality	Was the result as good or better?	Errors, rework, review time, customer satisfaction
Business outcome	Did the change create economic value?	Margin, revenue, cycle time, retention, avoided cost

Time Data Connect AI Use to Operating Results

Time data establish baselines for measuring labor productivity changes across several operational relationships:

- Employees and labor cost – time worked data
- AI tool usage within workflows – not only token use costs, but employee time spent using AI
- Workflows within projects – requires time tracking at the workflow level
- Projects performed for customers (internal, external) – per-customer project time worked
- Effort and output – time worked per unit of output

Let's say time tracking data show that AI reduced the time required to prepare a project report from 12 hours to 7. To know if that time reduction increased productivity and/or quality, you must ask:

- Did review time increase? Is it being tracked?
- Was the report's quality maintained or improved? Better quality achieved with the same labor effort is an effectiveness gain, even when no hours are released.
- Were the five released hours used productively? Have you decided how to identify which uses are productive uses in the time data?
- Did the organization complete more reports in the same amount of time?
- Did it improve margins or responsiveness?
- Was the improvement attributable to AI rather than another process change?

Four Ways that AI Generates Value

AI can create value in several ways beyond direct labor reduction. Four categories deserve particular investigation: efficiency, capacity, effectiveness, and acceleration. Controlled and field studies have found sizable productivity or quality gains in some writing, customer-support, and consulting tasks, but the effects vary by task, worker, and workflow. Measuring gains in any category requires suitable baseline data, with time data playing a central role when labor or workflow performance is involved.^{6,7,8}

Efficiency. AI creates an efficiency gain when the same work is completed with fewer labor hours. Time data can establish whether the reduction occurred. Further evidence is also needed to determine whether it produced value through lower delivery cost, improved project margins, reduced overtime, lower contractor spending, or productive redeployment. Efficiencies should be captured and redirected to achieve value in one of the other value categories.

Capacity. If time saved by AI is used to complete more work, organizations can increase capacity. Examples include more customers served, bigger projects undertaken (or more projects simultaneously), shorter backlogs, or faster growth without proportional hiring.

Effectiveness. Quality improvements unlocked by AI allow organizations to achieve better results without a corresponding increase in time and effort. Quantifying the value of effectiveness gains requires good historical data for time worked and other production costs to serve as a baseline. More effective organizations make fewer errors. Effectiveness improvements can also show up as higher conversion rates, improved customer retention, and reduced compliance or operational risk.

Acceleration. AI creates an acceleration gain when it shortens a consequential cycle or delivery period. Faster throughput creates financial value when it increases usable capacity, advances revenue or cash flow, reduces delay costs, or improves a consequential customer or operating outcome.

Each category requires different evidence. AI ROI cannot be established by forcing every use case into a simple “hours saved × hourly rate” formula.

The Measurement Framework

A practical AI ROI framework must account for the evidence needed at each measurement layer and for the different ways value may be realized. The following six steps provide a starting point for workflow-level measurement and evaluation.

Step 1: Define the appropriate units of work and cost

Break workflows down into critical work units, with regard to where a return on AI investment would deliver tangible value to your organization. It must be something specific enough to measure; otherwise gains can’t be tracked. Look at workflows such as preparing a proposal, reviewing a contract, producing a monthly report, resolving a support ticket, writing and testing a software feature, or completing an inspection analysis. “Knowledge work” or “marketing productivity” is too broad. Critical work units might include:

- Labor hours
- Elapsed time
- Output units
- Error or rework units or rate
- Review and approval effort

Workflow breakdown enables better effort and labor cost measurement. Direct costs are often broken down and measured by standard accounting workflows, and these cost breakdowns should be reviewed.

Step 2: Establish the pre-AI baseline

If reliable pre-AI data do not exist for the selected work units, establish a suitable comparison. Depending on the workflow, this might use historical records, a limited control period, matched work

groups, phased deployment, or selected work completed with and without AI. Avoid disrupting consequential work merely to reconstruct an artificial pre-AI state. Baseline evidence may come from time tracking, accounting, and other systems of record.

Step 3: Identify the AI interventions and understand the investment

AI, once introduced, may involve various tools used in different ways across the organization. Identify the specific AI interventions underway, including the tools used, workflow stage affected, roles involved, degree of human review, and training and implementation requirements. This inventory allows you to measure the “I” in your AI ROI. Your investment in AI is not only tool cost; it is also labor costs related to AI deployment, human review and rework of output, and training.

Step 4: Measure post-AI work and cost (including in the AI workflows)

Measure the complete AI-assisted workflow, including work introduced by the technology. These activities may become more efficient as employees gain experience, but they may also reveal costs or delays omitted from the original business case. AI-related work units could include:

- Prompt preparation
- Data preparation and retrieval
- Verification of AI output
- Editing output
- Reviewing output
- Rework
- Exception handling

Step 5: Compare post-AI measurements to baselines

Post-AI measurements from time tracking, accounting, and other systems allow a comparison against the pre-AI baselines. The comparison should be made over a time period that permits a credible result on which to base conclusions about the return on investment in AI. Thereafter, variances from the original baselines can be monitored for continuing returns, or new baselines can be set against which further improvements can be measured and judged.

Assuming comparisons reveal time savings through use of AI, discovering the true return on AI investment will require determining what happened to the hours released. It can also reveal ways to increase the return. Time saved might be redirected, for better or worse, to:

- Additional productive work
- Backlog reduction
- Customer-facing activity
- Innovation or improvement
- Idle or unmeasured time
- Elimination through attrition or reduced hiring

An hour released is an operational opportunity. It becomes financial value only when the organization uses or captures it deliberately.

Step 6: Test whether results persist at scale

Results from a contained pilot may not persist at enterprise scale. A larger deployment can introduce different work, users, case complexity, training requirements, integration costs, review burdens, and control risks. Measure those effects rather than extrapolating the pilot result directly.

Practical ROI Formulas

No single benefit calculation captures every way AI can create value. The following starter formulas can be used for the relevant value categories and then combined carefully.

Efficiency value:

$$(Baseline\ hours - AI-assisted\ hours) \times loaded\ labor\ rate$$

Note that this represents potential labor value, and not necessarily realized value. Realized value requires that efficiencies be redirected to achieve one of the other value categories.

Capacity value:

$$Additional\ units\ completed \times contribution\ value\ per\ unit$$

Quality value:

$$Reduction\ in\ errors\ or\ rework \times average\ cost\ per\ error\ or\ rework\ occurrence$$

Acceleration value:

$$(Baseline\ cycle\ time - AI-assisted\ cycle\ time) \times value\ of\ earlier\ completion\ per\ unit\ of\ time$$

Acceleration can overlap with efficiency, capacity, or quality. A shorter cycle time has monetary value only when it changes revenue, cost, cash flow, risk, customer outcomes, or usable capacity. Count the resulting economic benefit once, under the category that best represents how the value was realized.

Net AI value:

$$Total\ quantified\ benefits - Total\ AI\ costs$$

Total costs should include licenses, integration, training, governance, review, security, change management, and other material operating costs, not merely the subscription price.

AI return on investment:

$$(Total\ quantified\ benefits - Total\ AI\ costs) \div Total\ AI\ costs \times 100$$

ROI is expressed as a percentage. All the other category results are dollar amounts, not rates. A rate or unit value is used to convert an operational improvement into dollars. The “unit” might be an hour, transaction, project, error, customer, day, or some other measurable quantity. The crucial discipline is to label the period. “\$500,000 in net AI value” and “200% ROI” are incomplete unless the organization knows whether those figures cover one year, three years, or some other interval.

Include each economic value only once. Efficiency, capacity, quality, and acceleration should not be added together when two categories describe the same underlying value. For example, if efficiency gains are redirected to create greater capacity, use only the measured capacity value in the ROI calculation.

Attribution and Measurement Hazards

AI ROI measurements can appear more conclusive than the underlying evidence supports. Organizations should account for several common sources of distortion.

Self-reporting. Employees may overestimate or underestimate time savings or other value. Whenever possible, compare perceptions with observed time, output, and workflow data; one randomized developer study found that participants expected AI to make them faster even though measured completion time increased.⁹

Varying skill levels. AI may benefit novice or lower-performing employees more than experienced employees for some tasks, while other tasks may show a different pattern. A workforce-wide average can conceal these differences.¹⁰

AI learning curves. Early measurement may understate value while employees learn the tools and the organization redesigns work around them. Initial training, integration, and process changes can also temporarily increase measured cost.¹¹

Simultaneous process improvements. Other improvements may be introduced at the same time as AI; their effects should not be attributed entirely to the AI introduction.

Increased rework. AI may accelerate a first draft or prototype while increasing the time required for verification, correction, testing, or approval. Measure the complete workflow because AI performance can vary sharply even among apparently similar tasks.

Quality that is difficult to quantify. Improvements such as clearer analysis, better ideas, or more consistent customer communication are hard to measure. The measurement challenge is to define observable quality indicators whenever quality is critical to the ROI analysis.

Short measurement periods. Value may appear later. A short measurement period may miss these effects, and will also make conclusions drawn from the dataset less reliable.

Double-counting benefits. The same improvement should not be counted as both time savings and increased capacity if those measures describe the same released hours. Assign the benefit only once to the category of value most important to your operations.

Counting potential value as actual value. Saving three employee hours does not reduce a salaried employee's cost by three hours. The released time creates potential value; it becomes realized value only when it supports additional useful work, avoids another expense, increases output, or permits an actual reduction in labor cost.

Employee mistrust. Employees who believe AI will result in higher workloads, reduced staffing, or closer surveillance may be reluctant to disclose AI-enabled efficiencies. Mistrust is more likely if organizations fail to explain how measurements will be used.

How Time Tracking Should Be Designed for AI Measurement

When labor or workflow performance is material to the return, collect enough time data to establish the baseline and measure the change. Employees do not need to account for every prompt or minute. Instead, focus on the time categories that matter to the economic analysis:

- Track time to projects, customers, activities, or work types already meaningful to the business.
- Keep time categories stable long enough to establish a baseline.
- Add limited identifiers to tag AI-assisted workflows.
- Compare similar work across time periods or groups.

- Combine time data with other data, such as direct costs from accounting, and data from other operational systems rather than forcing the time system to contain every outcome metric.
- Explain that the purpose is process improvement and investment evaluation, and not employee surveillance.

The Executive Scorecard for AI Decisions

Ultimately, ROI from AI becomes an input to decisions about whether and how to integrate AI into operations. A concise scorecard for each significant AI use case could assist in summarizing the decision criteria. It could include:

- Business objective
- Workflow affected
- Population using AI
- Baseline labor hours
- AI-assisted labor hours
- Output or throughput change
- Quality or rework change
- Cycle-time change
- Released-time disposition
- Financial benefit
- Fully loaded AI cost
- Confidence rating
- Decision: stop, refine, maintain, or scale

Including a confidence rating prevents uncertain estimates from appearing more precise than the evidence supports. It can also justify further investigation of the ROI, depending on what is driving the uncertainty.

From AI Use to Measurable Value

Evidence shows that AI can make some work faster or better. The consequential question is whether those local improvements create measurable value for the enterprise. Organizations that establish credible baselines, measure labor effort across complete workflows, and connect changes in time to output, quality, capacity, and financial results can replace broad productivity claims with defensible evidence. When labor and workflow performance are material to the expected return, time tracking provides the measurement layer that connects AI-assisted work to economic value.

About the Author:

Amanda Finch is Director of Risk and Compliance at Journyx, where her work spans AI governance, compliance, risk management, and the practical use of workforce data to understand organizational performance and technology value.

Author's Note on AI Assistance:

The thesis, analysis, judgments, and original drafts of this paper are the author's own. Generative AI tools assisted with outlining, editorial refinement, and identification of potential sources. The author reviewed the final text, verified the cited sources, and takes responsibility for the paper's conclusions.

1 McKinsey & Company, "The State of AI: How Organizations Are Rewiring to Capture Value" (2025), <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>.

2 Stanford Institute for Human-Centered Artificial Intelligence, "Economy," The 2026 AI Index Report (2026), <https://hai.stanford.edu/ai-index/2026-ai-index-report/economy>.

3 Stanford Institute for Human-Centered Artificial Intelligence, "Economy," The 2026 AI Index Report (2026), reporting that AI-agent deployment remained in the single digits across nearly all surveyed business functions, <https://hai.stanford.edu/ai-index/2026-ai-index-report/economy>.

4 National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1 (2023), <https://doi.org/10.6028/NIST.AI.100-1>; see also NIST AI Resource Center, "Measure and Document Human Oversight of AI Systems," https://airc.nist.gov/AI_RM_F_Knowledge_Base/Playbook/Measure.

5 U.S. Bureau of Labor Statistics, "Labor Productivity and Total Factor Productivity," January 13, 2025, <https://www.bls.gov/productivity/educational-material/labor-productivity-total-factor-productivity-comparison.htm>.

6 Flavio Calvino, Julie Reijerink, and Lea Samek, "The Effects of Generative AI on Productivity, Innovation and Entrepreneurship," OECD Artificial Intelligence Papers, no. 39 (2025), <https://doi.org/10.1787/b21df222-en>.

7 Fabrizio Dell'Acqua et al., "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality," Organization Science, published online March 11, 2026, <https://doi.org/10.1287/orsc.2025.21838>.

8 Shakked Noy and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," Science 381, no. 6654 (2023): 187–192, <https://doi.org/10.1126/science.adh2586>.

9 METR, "Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity," July 10, 2025, <https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/>.

10 Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond, "Generative AI at Work," Quarterly Journal of Economics 140, no. 2 (2025): 889–942; originally NBER Working Paper 31161, <https://doi.org/10.3386/w31161>.

11 Erik Brynjolfsson, Daniel Rock, and Chad Syverson, "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," NBER Working Paper 25148 (2018), <https://doi.org/10.3386/w25148>.