THE AI INSTITUTE / OPERATING INTELLIGENCE

EVIDENCE PAPER · 02 / 2026

The Workflow Dividend

Why time saved is only inventory—and how organisations turn AI-enabled capacity into measurable enterprise value.

COOs, CFOs, transformation and functional leaders
Research period: January 2023 – July 2026
First published: 5 August 2026
Evidence current to 31 July 2026

PUBLICATION RECORD

An Institute thesis, supported by evidence

AI creates a workflow dividend only when changed work, released capacity and outcome economics are designed together. Time saved by an individual is not enterprise value until the organisation decides where that capacity goes.

WHAT LEADERS SHOULD TAKE FROM THIS PAPER

01	**Every AI investment needs a workflow theory of value, not a generic productivity claim.**
02	**Measure throughput, quality, rework and capacity allocation together.**
03	**Expect heterogeneous results by task, skill and operating context.**
04	**Promote use cases through evidence gates and stop those that cannot show durable economics.**

Research method

This paper synthesises peer-reviewed field evidence, randomised workplace studies, specialist task trials and OECD productivity analysis. The studies are treated as boundary cases rather than pooled into a universal productivity estimate. The Institute framework distinguishes observed findings from its operating inference.

First published 5 August 2026 · Updated 5 August 2026 · Research period January 2023 – July 2026 · Version 1.1 · Suggested citation: The AI Institute, *The Workflow Dividend* (2026).

CONTENTS

Inside this paper

How to read the evidence

External research supports the Institute's thesis; it does not substitute for it. Figures retain their original populations and definitions. Institute frameworks and inferences are labelled, and limitations are recorded before the references.

01 EXECUTIVE BRIEF

The missing step between speed and value

Generative AI can help people complete some tasks faster. That finding is useful, but incomplete. The organisation does not receive a dividend merely because an employee saves minutes. Value appears only when the surrounding workflow changes, released capacity is deliberately reallocated, quality remains inside limits, and the resulting outcome exceeds the full cost of deployment and operation.

The distinction explains why credible studies can point in different directions. A deployment across 5,172 customer-support agents increased issues resolved per hour by 15% on average, with the largest gains among less experienced workers. A six-month randomised field experiment involving 7,137 knowledge workers found active users spent about two fewer hours on email each week, yet detected no broader change in task quantity or composition from individual tool provision. A specialised trial found experienced open-source developers took longer with early-2025 AI tools despite believing they were faster. [1][2][3]

These results do not cancel one another. They show that value is conditional on the task, worker, tool, workflow and management response—a pattern also visible across Stanford's current synthesis of enterprise evidence. The Institute calls the realised gain the workflow dividend: the durable outcome produced after speed, quality, rework, risk, cost and capacity allocation are accounted for. [6]

> **Institute thesis — Time saved is a resource. The workflow dividend is the return created when that resource is deliberately converted into an outcome.**

02 WHAT THE EVIDENCE ESTABLISHES

AI productivity is real, uneven and local

The customer-support study offers strong evidence that AI can raise throughput in a bounded, measurable workflow. It also shows why averages mislead: less experienced agents improved most, while the highest-skilled workers saw smaller speed gains and some quality deterioration. The tool appeared to diffuse practices embodied in high-performing conversations, making the value mechanism partly one of expertise transfer. [1]

The knowledge-worker experiment shows a different boundary. Giving individuals a tool changed a narrow activity—email time—but did not automatically reorganise work. This is precisely what a workflow lens predicts. Individual assistance can create a pocket of capacity while meetings, approvals, hand-offs, service expectations and staffing remain unchanged. [2]

The developer trial is a warning against perceived productivity as the sole measure. Participants predicted that AI would make them faster and continued to believe it had, while measured completion time was 19% longer. The sample was small and unusually experienced, and later tools may perform differently. The finding remains important because it exposes the gap between felt fluency and observed outcome. [3][4]

OECD's 2026 productivity synthesis cites firm-level analysis associating AI use with a short-run labour-productivity increase, but also stresses that adoption definitions are not harmonised, causality is difficult and complementary intangible investment matters. Its broader adoption research likewise treats skills, data, finance, infrastructure and management capability as complements rather than background conditions. The responsible conclusion is not a universal return estimate. It is that leaders must construct local evidence at the level where work and outcomes can actually be observed. [5][7]

Provider-native customer telemetry can add useful evidence about usage growth and task mix, but it describes a selected customer population and should not be converted into a universal productivity or return estimate. It is an operating input to investigate, not proof that value has reached the enterprise. [8]

03 INSTITUTE FRAMEWORK

A workflow theory of value

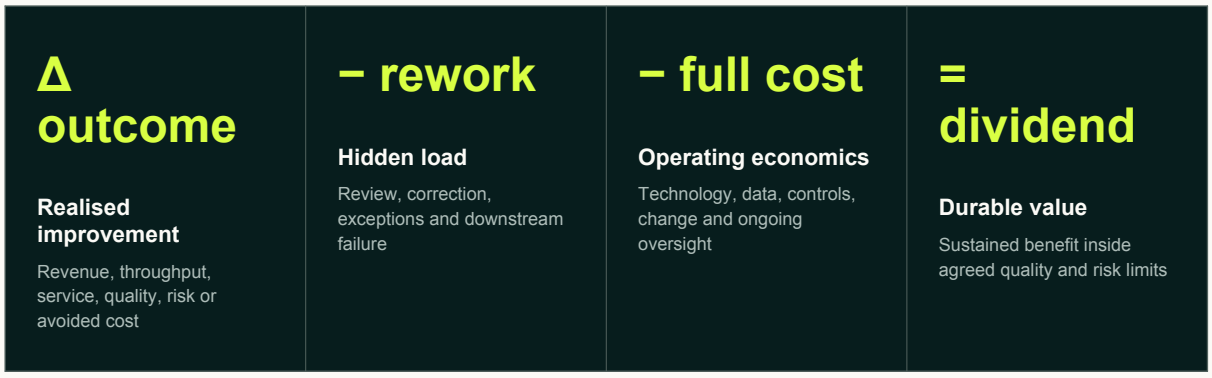
Before funding a use case, write a one-page theory that links the capability to an enterprise outcome. Start with the baseline: volume, cycle time, quality, cost, risk and current variation. Describe the changed work: which tasks disappear, accelerate or become possible; which hand-offs and decisions change; and what people must do differently.

Then specify the capacity destination. Saved time may be absorbed by higher service levels, more volume, faster revenue, better decisions, reduced contractor spend, fewer errors or a smaller future cost base. Each pathway needs a different measure. If the organisation cannot name the destination, time saved is likely to disappear into work expansion, queues, meetings or unmeasured discretionary effort.

Finally define quality and risk guardrails, full costs and a stop rule. Costs include licences, integration, data preparation, evaluation, human review, change support, monitoring, incidents and model or process maintenance. A use case that looks attractive on inference cost alone can become uneconomic once review and exception handling are visible.

The workflow dividend is therefore not a model benchmark. It is the measured change in the target outcome, adjusted for quality, risk and full operating cost, that can credibly be attributed to the changed workflow.

EXHIBIT / THE WORKFLOW DIVIDEND



04 EVALUATION DESIGN

Measure the work, not the enthusiasm

A credible evaluation begins before deployment. Establish a baseline using the same definitions, population and observation window that will be used after the change. Where possible, use phased rollout, a comparable control group or within-workflow randomisation. Where that is impractical, use a stable interrupted time series and document other changes that could explain the result.

Track a balanced set of measures: outcome, throughput, cycle time, first-pass quality, rework, exception rate, customer or employee impact, control failures, unit cost and total capacity used. Segment results by task and worker experience. An average improvement can conceal harm to expert work, exclusion of difficult cases or a transfer of effort to reviewers and downstream teams.

Separate leading from lagging evidence. Usage, satisfaction and self-reported time saved can guide adoption support. They should not approve scale. Promotion requires observable workflow behaviour and outcome evidence. Finance should validate value claims, while risk and operational owners validate quality and control evidence.

Re-measure after the novelty period. Workarounds, prompt habits, vendor changes, model drift and changing task mix can all alter economics. A workflow that cleared its initial gate should remain subject to periodic value and control review.

05

PORTFOLIO MANAGEMENT

Fund evidence, not use-case volume

Treat AI initiatives as an option portfolio. Early experiments should be cheap and designed to reduce a named uncertainty: technical feasibility, data availability, user behaviour, quality, control effectiveness or unit economics. Money and authority increase only as evidence strengthens.

Use four gates. Discovery confirms the workflow and baseline. Trial tests whether the changed work improves a proximal measure. Controlled production establishes quality, controls and full operating cost under real conditions. Scale requires durable outcome evidence, an operating owner and a capacity-allocation decision.

A stop decision is a sign of portfolio health. Terminate work when the task is too variable, review costs erase the benefit, the workflow cannot absorb saved capacity, adoption requires disproportionate change, or a better non-AI redesign exists. Keeping weak pilots alive creates maintenance debt and obscures the few investments with enterprise potential.

Report portfolio quality through evidence movement: how many hypotheses were tested, what uncertainty was retired, which initiatives advanced or stopped, what value was validated, and where benefits failed to reach the enterprise.

EXHIBIT / FOUR EVIDENCE GATES

01	02	03	04
Discover	**Trial**	**Control**	**Scale**
Name the workflow, outcome, baseline and highest uncertainty.	Test changed work and proximal performance with a bounded cohort.	Expose quality, rework, exceptions, risk and full operating cost.	Validate durable outcomes and deliberately allocate released capacity.

06 OPERATING ECONOMICS

The destination of saved capacity determines the dividend

Time savings become value through a management decision. Capacity can absorb unmet demand, improve service, reduce overtime, increase throughput, support training or lower required resources over time. If no destination is chosen, saved minutes fragment across the day and disappear from enterprise accounting.

The destination changes the evidence required. A service-level claim needs queue and response measures. A quality claim needs defects, rework and customer outcomes. A staffing claim needs sustainable workload evidence and a transition plan. A learning claim needs skill progression rather than self-reported confidence. Australia's productivity analysis similarly treats realised gains as conditional on adoption, complementary investment and policy settings. [10]

Team-level experimental evidence suggests AI can alter how expertise is distributed and can enable an individual with AI to match some team outcomes in a product-innovation setting. That is a provocative configuration result, not a general instruction to remove teams. The social, coordination and development functions of teams remain part of the operating decision. [9]

Finance and operations should therefore agree a capacity ledger before scale: baseline hours and demand, the proposed destination, the mechanism that makes capacity fungible, expected lag, quality guardrails and the owner who will realise the change.

EXHIBIT / FIVE DESTINATIONS FOR RELEASED CAPACITY

Serve	Improve	Learn	Reduce
Absorb demand	**Raise quality**	**Build capability**	**Change the cost base**
More volume or shorter queues without proportional resource growth.	Use capacity for review, personalisation, prevention or complex cases.	Protect practice, coaching and judgment development.	Capture sustainable resource change after demand and quality evidence.

07 WORKED EVALUATION

Measure the human–AI workflow, not the demo

Begin with representative cases sampled across routine work, exceptions and high-consequence edge conditions. Record baseline cycle time, throughput, quality, rework and hand-offs. Define the minimum worthwhile improvement and the quality floor before teams see treatment results.

Where feasible, compare current work, assisted work and a redesigned workflow. Segment by experience because average effects can hide who benefits and who becomes less reliable. The HBS frontier experiment and later customer-support evidence both show why task boundary and worker heterogeneity matter. [1][12]

Move from trial to operation through an evidence window, not a permanent approval. Continue measuring review burden, exception rate, incident signals, unit cost and the destination of saved capacity. Treat a model, prompt, data, integration or population change as a reason to reassess transportability. NIST's generative-AI profile supports lifecycle testing and monitoring calibrated to the deployed context. [11]

The stop rule should be as specific as the scale rule: terminate or redesign when quality falls below the floor, review consumes the gain, the workflow cannot absorb capacity, risk cannot be bounded or a simpler process change creates a better outcome.

08 90-DAY AGENDA

Make the value mechanism explicit

In the first 30 days, choose three material AI-enabled workflows and reconstruct their theory of value. Replace claimed hours saved with a baseline, capacity destination, balanced measures and full-cost estimate. Identify where value currently leaks into review, exceptions or unchanged operating constraints.

By day 60, establish common evidence gates and require a finance, operational and risk sign-off appropriate to each stage. Instrument the chosen workflows so throughput, quality, rework and unit cost can be observed together. Define what result would cause the organisation to stop.

By day 90, make one explicit scale decision, one redesign decision and one stop decision. Reallocate funding based on evidence quality and expected workflow dividend, then publish the assumptions so the executive team can revisit them as models, costs and work change.

RESEARCH RECORD

Method and limitations

Method

This paper synthesises peer-reviewed field evidence, randomised workplace studies, specialist task trials and OECD productivity analysis. The studies are treated as boundary cases rather than pooled into a universal productivity estimate. The Institute framework distinguishes observed findings from its operating inference.

Limitations

The evidence base is moving quickly and remains concentrated in selected tools, firms, countries and knowledge-work tasks. Tool capability changes faster than publication cycles. Local evaluation can still be confounded by selection, learning effects, simultaneous process changes and unmeasured quality transfer.

Evidence conventions

Source facts are cited to the original publication. Institute analysis reconciles definitions or combines evidence. Institute inference extends that analysis into a management proposition. Frameworks are decision aids, not claims of universal causality. Every important statistic should be read with its population, date, definition and caveat.

Research cutoff: 31 July 2026. Review cycle: six months, or earlier after a material change in evidence, standards or policy.

REFERENCES

Evidence behind the thesis

01 **Brynjolfsson, Li and Raymond, Generative AI at Work, Quarterly Journal of Economics**

Staggered deployment across 5,172 customer-support agents.

02 **Dillon et al., Shifting Work Patterns with Generative AI, NBER Working Paper 33795**

Six-month randomised field experiment covering 7,137 knowledge workers at 66 firms.

03 **METR, Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity**

Randomised trial across 246 tasks; small, specialised sample.

04 **METR, Uplift Update**

Follow-up documenting selection and time-measurement limits.

05 **OECD, Compendium of Productivity Indicators 2026**

Cross-country synthesis; warns about harmonisation and causal limits.

06 **Stanford HAI, 2026 AI Index — Economy**

Current synthesis of adoption, investment and productivity evidence.

07 **OECD, The Adoption of Artificial Intelligence in Firms**

Evidence on complementary investment and adoption barriers.

08 **OpenAI, The State of Enterprise AI 2025**

Provider-native customer usage and survey data; not representative of all firms.

09 **Dell'Acqua et al., The Cybernetic Teammate**

Preregistered field experiment with 776 professionals examining performance, expertise sharing and teamwork.

10 **Australian Productivity Commission, Harnessing Data and Digital Technology**

Australian productivity modelling and policy analysis with explicit uncertainty.

11 **NIST AI 600-1, Generative AI Profile**

Testing and risk-management actions for deployed generative-AI systems.

12 **Dell'Acqua et al., Navigating the Jagged Technological Frontier**

Preregistered field experiment demonstrating task-boundary and user-configuration effects.

THE AI INSTITUTE

Evidence should change what you do next.

Research · Readiness · Capability

theai.institute