

THE

AI ADOPTION OPERATING REPORT

From access to accepted work: evidence, economics, and a 90-day plan
for leaders



EXECUTIVE BRIEF

The goal is useful work, not more seats

A focused synthesis of public research and an original operating method - not a new survey.

Buying a tool makes AI available. It does not tell you whether a customer problem was solved, a decision improved, or a worker got back time they could actually use. This report starts with the completed job and works backward to the source, review, permission, and economics needed to make it repeatable.

U.S. FIRMS

18%

Used AI in a business function in Nov 2025-Jan 2026. Census working paper [1].

LEADERS SURVEYED

25%

Had moved at least 40% of experiments into production. Deloitte [2].

JOBS NOT REDESIGNED

84%

Deloitte respondents reporting no job redesign around AI [2].

Four measures worth keeping separate

1

Access

Who has an approved tool and what work are they allowed to do?

2

Activity

How many eligible tasks actually used it, and where did people stop?

3

Accepted work

How much output survived review without harmful corrections?

4

Business result

Did service, quality, cycle time, margin, or customer experience improve?

These are different studies, populations, and questions. Do not read 18%, 25%, and 84% as stages in one funnel.

HOW TO USE THIS REPORT

What this report can - and cannot - prove

A decision guide built from distinct forms of evidence.

We reviewed public primary sources available through October 10, 2026: the 2026 U.S. Census working paper, Deloitte's 2026 enterprise survey, field experiments, a study of 51 selected successful deployments, and NIST's risk framework. We did not run a new survey or test a vendor product for this report. The scorecards and financial examples that follow are our operating tools, not validated industry benchmarks.

Read each evidence type for the job it can do

Representative firm data [1]

Describes how broadly U.S. firms reported AI use under a specific question. It does not measure whether use paid off.

Executive survey [2]

Shows what 3,235 senior leaders in 24 countries reported in Aug-Sep 2025. Expectations about future scale are not realized outcomes.

Workplace experiments [3, 4, 7, 8]

Estimate effects in particular tasks, tools, and organizations. A result in support or coding is not your firm's expected return.

Selected deployment cases [5]

Show concrete implementation work. Stanford's 51 cases were chosen for success and cannot tell us the success rate of all AI projects.

Risk guidance [6]

NIST supplies voluntary practices for mapping, measuring, managing, and governing risk. It does not certify a given use case as safe.

Use this as a pilot design and review aid. Replace every example threshold with a risk limit approved for your own work.

THE ADOPTION PICTURE

Three snapshots, three different questions

A useful chart preserves the population and question behind each number.

**18%****U.S. FIRMS****AI used in a business function**

Nov 2025-Jan 2026 reference period; Census working paper [1].

25%**DELOITTE RESPONDENTS****40%+ experiments moved to production**

3,235 leaders surveyed Aug-Sep 2025 across 24 countries [2].

84%**DELOITTE RESPONDENTS****No jobs or nature of work redesigned**

Self-reported organizational practice, not a measured productivity effect [2].

What to take from this

There is plenty of room to improve. The challenge is not simply awareness: access, production deployment, and redesigned work are separate problems. A company can have a high seat count and still have no reliable path from a model's answer to an accepted decision or customer outcome.

The Census paper changed its core AI-use question in November 2025. Its 18% figure must not be stacked on the site's earlier 5.4% measure of firms using AI to produce goods or services. Deloitte's 25% is a threshold among a different, international leader sample. The figure above deliberately avoids a misleading trend line or funnel [1, 2].

WHAT MEASURED STUDIES FOUND

AI can help - but the job matters

Three original studies, three different outcomes and populations.

CUSTOMER SUPPORT

+14%

Issues resolved per hour

A staggered rollout to 5,179 support agents found an average gain; less experienced and lower-skilled agents saw about 34% [3].

LIMIT: One software company's support workflow; not a general call-center forecast.

KNOWLEDGE WORK

1.4 h

Less email time per week

In a randomized six-month study of 6,000 workers, access to an AI assistant reduced weekly email time by 1.4 hours on an intent-to-treat basis [4].

LIMIT: Among active users the reported difference was 3 hours; meeting time did not materially change.

SOFTWARE DEVELOPMENT

+26%

Completed tasks

Three randomized company experiments covering 4,867 developers found a combined 26.08% increase in completed tasks [8].

LIMIT: The combined estimate had a 10.3-percentage-point standard error; results varied by setting.

The comparable question for your company is not 'Will we get 14%?' It is 'Which accepted outcome can we measure against our own baseline?'

CASE STUDY

A logistics invoice project needed a new process

Stanford's selected-case report is useful for the work behind the result, not as a population average.

A U.S. refrigerated-trailer company in Stanford's 2026 Enterprise AI Playbook handled more than 100,000 maintenance invoices each year. Seven full-time-equivalent roles worked through inconsistent formats and handoffs. The published case describes an eight-week implementation with 85% accuracy, processing in less than 24 hours, and a shift from seven to two full-time equivalents in this workflow [5]. Those are participant-reported case figures from an anonymized, success-selected sample.

The less photogenic work made the system possible

1

Simplify the source

The team reduced a sprawling set of redundant invoice templates before asking the model to map them.

2

Make experts available

Subject-matter specialists reviewed thousands of outputs and explained errors alongside their normal jobs.

3

Give operations the ownership

The president checked bottlenecks weekly, while junior IT staff learned to operate the system rather than inherit a black box.

4

Keep an exception route

The reported 85% accuracy still leaves work that needs identification, correction, and accountability.

Our inference: count template cleanup, expert review, integration, and error handling in the business case. The model's license is not the deployment cost.

In the same selected set of 51 successful cases, 61% had a prior failed AI effort [5]. That is not an industry failure rate; it shows why the cost of learning should be visible in a budget.

CHOOSE THE FIRST JOB

Start where the work is visible

The best first use case is important enough to matter and narrow enough to judge.

Pick a recurring workflow, not a department. 'Use AI in marketing' is too vague to measure. 'Draft the weekly campaign performance note from approved reports, with a named reviewer before distribution' has an owner, input, comparison, and finish line.

Five filters before a pilot begins

1

An observable outcome

Can the team name a finished case, document, decision, or customer action?

2

A baseline

Can you retrieve 20-30 recent examples and see current time, corrections, and failures?

3

Accessible source material

Does the system have approved, current records to use - and a way to cite them?

4

A willing owner

Will someone review real outputs, record errors, and make the scale/stop decision?

5

A safe first boundary

Can the trial begin without sending messages, moving money, or changing records automatically?

Poor first bets: a rare high-stakes decision, a workflow nobody owns, or a process with no reliable source records. Fix the work before adding a model.

USE-CASE PORTFOLIO

Four jobs a leader can actually test

Constructed examples: adapt the sources, owner, and stop rule to your organization.

CUSTOMER SERVICE**Answer a routine policy question from approved material**

FIRST BOUNDARY Show a source-linked draft to an agent; the agent sends it after checking.

MEASURE Correct resolutions per eligible request; repeat contact.

STOP IF A stale policy is cited or an exception is treated as routine.

FINANCE**Prepare an invoice exception for review**

FIRST BOUNDARY Extract fields and show mismatches; do not approve payment or change the ledger.

MEASURE Accepted exceptions cleared per reviewer hour; correction rate.

STOP IF A supplier or amount is matched to the wrong record.

MARKETING**Produce a campaign brief from approved results**

FIRST BOUNDARY Draft claims and variants; a campaign owner approves each factual claim.

MEASURE Approved brief cycle time and tested campaign outcome.

STOP IF AI turns an inferred trend into an unsupported guarantee.

OPERATIONS**Maintain a daily open-issue summary**

FIRST BOUNDARY Propose updates from tickets and logs; an owner marks what is resolved.

MEASURE Important issues found, missed issues, and repair time.

STOP IF A closed ticket is mistaken for a fixed customer problem.

OPERATING DESIGN

Build a path from evidence to accepted work

A source, a model, and a reviewer each have a different job.



AI-generated editorial illustration: the work must reach a real outcome, not circle inside a pilot.

1 / State the purpose

A human owner writes the problem, what a correct result must contain, and what the system must never infer.

2 / Bound the evidence

Name the approved records, their update schedule, and what happens when sources conflict or go stale.

3 / Draft and show the trail

The AI proposes an answer or action and points to the records used. Unknowns stay visible.

4 / Review and finish

A person accepts, edits, or rejects the output. The final business record retains who decided and why.

CONTROLS THAT PEOPLE CAN USE

Make the approved route easier than the workaround

A control is effective only if it is clear, workable, and recoverable.



AI-generated editorial illustration: a useful permission gate includes a legitimate way through.

Start with read-only access to approved material. Keep publishing, payments, customer messages, and record changes behind separate permissions. Record the prompt, sources, proposed action, reviewer decision, and repair path for significant cases. NIST's voluntary framework points organizations toward mapping the context, measuring risk, managing it, and assigning governance [6].

Avoid adding ritual steps nobody can complete. In a 2026 experiment with 388 retail employees, a rigid paired-use AI protocol was associated with lower document quality and much less output than unstructured use; the authors also report design limitations [7]. That does not prove all controls reduce quality. It is a warning to test whether a control improves the work instead of simply looking cautious.

Ask users where the approved path is too slow. Repair the path before a workaround becomes normal practice.

PEOPLE AND TRAINING

Teach the job, then teach the tool

Access alone leaves workers to invent their own habits and review standards.

Training should begin with a familiar piece of work. Give people a good past case, an awkward case, and an approved source. Show how to ask for a draft, inspect where each claim came from, correct a wrong answer, and decide whether the output is good enough to send. The review standard belongs to the business team, not only to IT.

A four-part practice session

1

Demonstrate

The process owner works through one real case and explains why the final answer is accepted or rejected.

2

Try independently

Each worker repeats the task with a different case, records the AI's contribution, and marks uncertainty.

3

Compare judgments

Reviewers compare outputs, not prompt cleverness. Differences reveal unclear policy or weak source records.

4

Repair the workflow

Users report slow approvals, missing records, and tempting workarounds. The owner fixes one point each week.

What the studies suggest

The customer-support study found much larger gains for novices than for experienced workers [3]. That makes training and skill transfer a plausible adoption goal, not a guaranteed outcome. A separate 2026 field experiment with 388 retail employees found that one rigid paired-use protocol hurt document production and quality in that setting [7]. Test your own training design against accepted work, not attendance.

Do not equate use with competence. Ask each team to maintain a small library of accepted examples and mistakes worth avoiding.

MEASUREMENT

Count the whole job, including correction

A pilot scoreboard should distinguish speed, quality, use, and value.

MEASURE	DEFINITION	WHY IT MATTERS
Eligible work	Number of cases that could use AI	Prevents a pilot from counting only easy cases
Accepted work	Share approved without material correction	Measures usable output, not draft volume
Review load	Minutes of checking and repair per case	Catches the bottleneck moved to a reviewer
Outcome	Resolved demand, conversion, cycle time, or errors	Ties the test to the business job
Trust and harm	Wrong-source, privacy, and customer incidents	Creates a reason to pause expansion
Net cost	Tools + setup + review + training + incidents	Stops a cheap model from hiding expensive work

Set the baseline before the new tool is visible. Sample hard cases as well as easy ones. When possible, compare similar cases or teams rather than a before/after period that also changed staffing, demand, or policy. Report the number of eligible cases next to every rate.

Sources on this page: [3, 4, 8] Studies show task-specific effects; measures below are editorial

ECONOMICS

Released capacity is not cash savings

A useful business case keeps four value streams separate.

Suppose a team handles 2,000 eligible cases a month. AI saves four drafting minutes on 60% of them, but adds one minute of review to each AI-assisted case. Net time released is 60 hours a month: $2,000 \times 60\% \times (4 - 1) / 60$. This is an illustrative calculation, not a reported company result.

CAPACITY RELEASED

60 hours

Time that can be assigned to other work, if the workflow and demand allow it.

CASH SAVINGS

\$0 until acted on

A budget changes only with avoided spend, overtime, vacancies, or a deliberate staffing decision.

A financial review should show each line

CAPACITY	Hours genuinely freed after checking and correction. Name the work that will absorb them.
CASH	Avoided or removed spend documented by Finance. Do not multiply hours by salary and call it savings.
GROWTH	Incremental revenue or margin from an attributable customer or sales effect - not the total value of assisted deals.
RISK AND QUALITY	Errors avoided, response times improved, or customer harm prevented; keep these separate from cash unless valued with a defensible method.

Subtract licenses, integration, source cleanup, training, reviewer time, monitoring, and recovery. Do not add overlapping benefits twice.

PILOT PLAN / DAYS 1-30

Run real work before asking for scale

Four weeks, one owner, a visible comparison, and an explicit pause rule.

WEEK 1**Describe the old way**

Collect 20-30 recent cases, including difficult ones. Record time, corrections, outcome, and exceptions. Name the process owner and reviewer.

OUTPUT: Baseline sheet

WEEK 2**Test without acting**

Connect only approved sources. Ask AI to propose outputs beside the old process. Review every factual claim and keep any external action manual.

OUTPUT: Error log

WEEK 3**Try with limited live work**

Allow a narrow set of low-risk cases. Record accept, edit, reject, review time, wrong sources, and user workarounds.

OUTPUT: Side-by-side results

WEEK 4**Decide what earned trust**

Compare accepted outcomes and total cost with the baseline. Examine failures individually. Expand only if the benefit remains after correction.

OUTPUT: Go / revise / stop memo

SCALE PLAN / DAYS 31-90

Grow the source set before autonomy

Expansion is a sequence of earned permissions, not a switch.

DAYS 31-45

Fix the source defects

Repair stale records, missing permissions, ambiguous definitions, and duplicate templates found in the pilot. Re-run the same difficult cases.

DAYS 46-60

Broaden cases, not actions

Add another eligible cohort or a higher volume of read-only cases. Keep the same review rule until errors are understood.

DAYS 61-75

Add one controlled action

If evidence supports it, allow one reversible internal update with a named approver, an audit record, and rollback.

DAYS 76-90

Review economics and ownership

Finance checks actual cost and cash effects; operations checks accepted work; security checks incidents and access; workers identify friction.

The scale decision

Do not expand because the demonstration looked polished. Expand when accepted work improves on repeated real cases, the reviewer workload stays manageable, the source trail is reconstructable, and the financial owner can state what value actually appeared. Otherwise revise the workflow or stop it.

WORKSHEET / READINESS

Score your own operating readiness

This 24-point rubric is an editorial tool, not a validated benchmark or certification.

For each dimension, choose 0 (missing), 1 (named), 2 (documented), 3 (tested on real cases), or 4 (measured and routinely improved). Record evidence, not optimism. A low score is a place to work, not a judgment on the company.

BUSINESS JOB

0 1 2 3 4

Is the outcome and process owner named?

SOURCES

0 1 2 3 4

Are approved records current and citable?

QUALITY

0 1 2 3 4

Do hard cases have a review standard and error log?

PEOPLE

0 1 2 3 4

Are users trained and able to report workarounds?

CONTROLS

0 1 2 3 4

Are permissions, incident repair, and rollback tested?

ECONOMICS

0 1 2 3 4

Are full cost and accepted value tracked separately?

Total ____ / 24 Lowest dimension: _____

Do not average away a critical zero. A missing owner, unusable source, or absent recovery route can stop a pilot regardless of the total.

WORKSHEET / ONE-PAGE CHARTER

Define one pilot in plain language

A completed charter should fit on one page and survive a skeptical review.

The recurring job and the person who owns it

The customer or business result that should improve

The cases included - and the cases explicitly excluded

Approved sources, their owners, and how freshness will be checked

What AI may propose; what it may never send or change

The reviewer, accept/reject rule, and recovery route

The baseline, comparison, and minimum useful improvement

The pause rule for wrong sources, privacy, harm, or excess review

The decision date and the people who can approve expansion

If the charter cannot name a completed unit of work, the project is still an idea, not a pilot.

DECISION GATE

Know when to expand, revise, or stop

Use your approved thresholds. The examples below illustrate the logic; they are not universal standards.

EXPAND

Accepted work improves

Repeatable gains in outcome or cycle time after review; source trail is clear; risk incidents are within agreed limits; owner and Finance accept the economics.

REVISE

The idea is good but the route is broken

Benefits appear only on easy cases, source data is stale, review consumes the saved time, or users bypass a slow approved workflow. Fix and retest.

STOP

The team cannot reconstruct or contain errors

Wrong sources recur, private information leaves the approved boundary, significant customer harm occurs, or no named owner will sign off.

A stop decision is useful research. Record the failure pattern so the next attempt starts with a better source, control, or work design.

RESEARCH NOTES

How to read the evidence responsibly

Sources support the finding stated beside them - not every broader claim about AI.

Census [1]

The 18% estimate concerns U.S. firms using AI in a business function in Nov 2025-Jan 2026. It is a research working paper, not an official Census publication. The question changed in late 2025, so earlier figures are not a simple trend.

Deloitte [2]

The survey sampled 3,235 director-to-C-suite leaders in 24 countries in Aug-Sep 2025. Percentages are respondent reports or expectations; they are not audited deployment outcomes.

Workplace studies [3, 4, 7, 8]

Effects are tied to the studied task, participant group, product, and time. Microsoft Research's 2026 collaboration experiment [7] reports design limitations; its result should prompt testing, not a blanket rule against structure.

Stanford cases [5]

The 51 cases were selected for documented value and rely partly on participant reports. The reported 61% with a prior failure is a share of selected successes, not a failure rate for all projects.

Our contribution

The four-measure ladder, economic separation, 30/60/90-day sequence, six-dimension scorecard, and pilot charter are AI Adoption Trends editorial frameworks. No claim is made that they have been externally validated.

SOURCE LIBRARY

Primary sources and next steps

Direct links are embedded in this PDF. Research verified October 10, 2026.

- [1] **Bonney et al., CES-WP-26-25 (2026), U.S. Census Bureau working paper**
<https://www2.census.gov/library/working-papers/2026/adrm/ces/CES-WP-26-25.pdf>
- [2] **Deloitte, State of AI in the Enterprise (2026); survey fielded Aug-Sep 2025**
<https://www.deloitte.com/content/dam/assets-zone3/us/en/docs/services/consulting/2026/state-of-ai-2026.pdf>
- [3] **Brynjolfsson, Li and Raymond, Generative AI at Work, NBER 31161**
https://www.nber.org/system/files/working_papers/w31161/w31161.pdf
- [4] **Dillon et al., Shifting Work Patterns with Generative AI (2025)**
<https://www.microsoft.com/en-us/research/publication/shifting-work-patterns-with-generative-ai/>
- [5] **Pereira, Graylin and Brynjolfsson, Enterprise AI Playbook (2026)**
https://digitaleconomy.stanford.edu/app/uploads/2026/03/EnterpriseAIPlaybook_PereiraGraylinBrynjolfsson.pdf
- [6] **NIST, AI Risk Management Framework and Generative AI Profile**
<https://www.nist.gov/itl/ai-risk-management-framework>
- [7] **Farach et al., Scaffolding Human-AI Collaboration (2026)**
<https://www.microsoft.com/en-us/research/publication/human-ai-collaboration-field-experiment/>
- [8] **Cui et al., Effects of Generative AI on High-Skilled Work (2026)**
<https://pubsonline.informs.org/doi/10.1287/mnsc.2025.00535>

Keep the decision moving

Use the one-page charter with the operations owner, the scorecard with the risk and data leads, and the four-week plan with the team that will perform the work. If the first trial is not measurable, narrow the job. If it is measurable but poor, fix the source, review, or permission step before buying more seats.

Further reading: aiadoptiontrends.com/tools/roi-calculator and aiadoptiontrends.com/training/ai-security