BIZNEWSFEED

# Generative AI at Work

*What the evidence says about productivity, quality, job exposure and responsible business use*

**Publication date: 12 September 2026**

Publisher: BizNewsFeed.com editorial team

**Editorial disclosure.** This whitepaper was prepared for BizNewsFeed as educational editorial content. It is not independent research commissioned by an unaffiliated third party, and it is not investment, legal, accounting, or other professional advice.

# Executive summary

Generative AI is diffusing quickly through business, but “AI adoption” is not the same thing as proven productivity. OECD data show that 20.2% of firms in countries with available data reported using AI in 2025, up from 8.7% in 2023.[1] Controlled and field studies show substantial gains on some bounded tasks: professional writing participants completed tasks 40% faster with an 18% increase in rated quality,[2] while a customer-support deployment was associated with a 14% average increase in issues resolved per hour and larger gains among less-experienced workers.[3] A consulting experiment likewise found strong gains on tasks inside the tested capability frontier, but worse performance on a task outside it.[4] The practical lesson is therefore conditional: firms should treat generative AI as a task-level capability that requires verification, governance and human accountability rather than as an across-the-board substitute for expertise.

# Intended audience and questions addressed

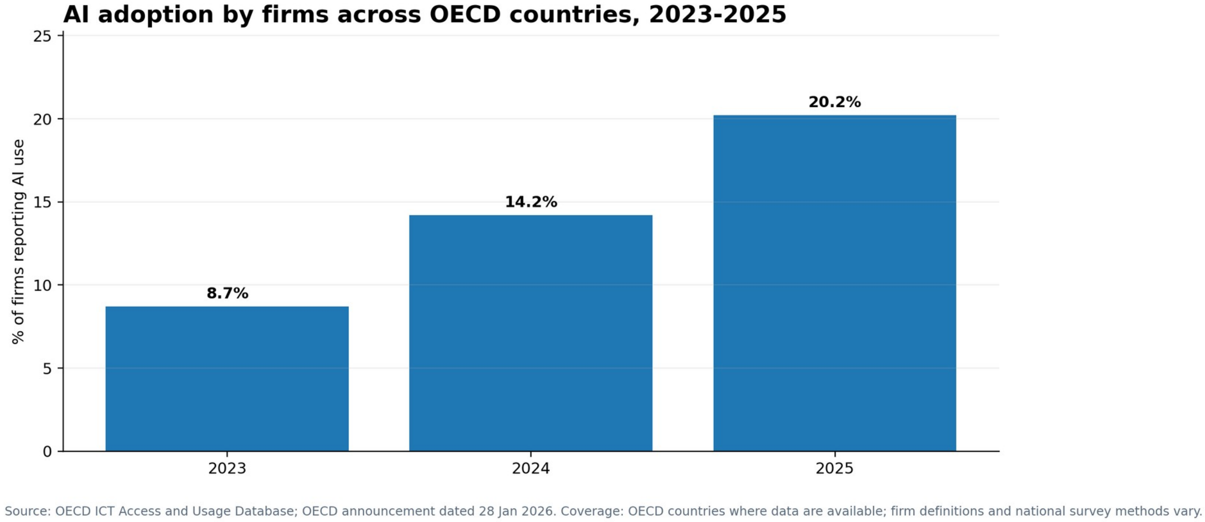
Business leaders, founders, managers, technology buyers and professionals evaluating where generative AI can create value without weakening reliability, privacy, security or accountability.

- Which productivity findings are supported by experiments, and what do they not prove?
- Which tasks are more or less suitable for generative AI assistance?
- How should a company structure human review, measurement and risk controls?
- What does job exposure mean, and why is exposure not equivalent to job loss?

# 1. Adoption is accelerating, but adoption is not an outcome

Firm-level AI use has risen rapidly. OECD reporting for 2025 shows 20.2% of firms using AI across countries where comparable data are available, versus 14.2% in 2024 and 8.7% in 2023.[1] Large firms reported much higher use than small firms, illustrating a diffusion gap rather than universal adoption. These statistics describe use, not value: a business can “use AI” in a pilot, a writing assistant or a production system, and those activities have very different economic and risk implications.

### Figure 1. AI adoption has more than doubled in two years




*Share of firms reporting AI use in OECD countries with available data, 2023-2025.*

Source: OECD [1].

**Takeaway: The adoption curve is steep, but the statistic does not tell us whether a particular implementation improves profit, quality or resilience.**

## 2. What controlled studies actually show

### Writing tasks: faster completion and higher rated quality

In a preregistered experiment with 453 college-educated professionals completing occupation-specific writing tasks, access to ChatGPT reduced average completion time by 40% and increased evaluator-rated quality by 18%.[2] The study is informative because assignment was randomized, but it covers a defined set of writing tasks under experimental conditions. It does not establish that the same gains occur in strategy, legal analysis, financial forecasting, software engineering or any other activity with different error costs and information requirements.

### Customer support: gains were uneven across workers

A field study of 5,179 customer-support agents found that access to a generative-AI assistant increased issues resolved per hour by 14% on average, with a 34% improvement among novice and lower-skilled workers and minimal effects among the most experienced workers.[3] This heterogeneity matters. AI can compress performance differences by making useful patterns easier to access, but experienced workers may already possess much of the tacit knowledge the system is surfacing.

### Knowledge work: the “jagged frontier”

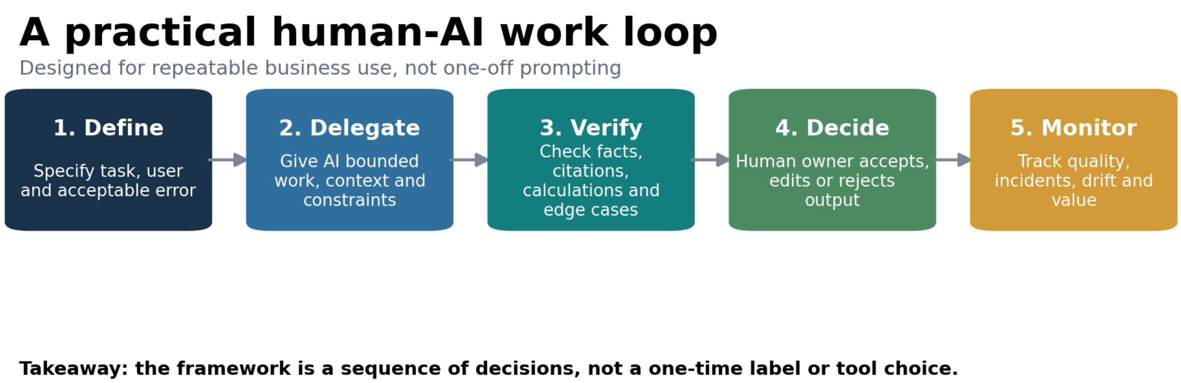
A field experiment involving 758 consultants found participants using GPT-4 completed more tasks, worked faster and achieved higher human-rated performance on tasks designed to fall within the model’s capability frontier.[4] On a task deliberately selected outside that frontier, AI users were less likely to reach the correct conclusion. The phrase “jagged technological frontier” captures a central business risk: model competence is not smooth. A system can be excellent at one adjacent task and unreliable at another that looks deceptively similar.

# 3. Job exposure is not the same as automation

The ILO's 2025 global task-level assessment estimates that roughly one in four workers are in occupations with some degree of generative-AI exposure, while 3.3% of global employment falls into the highest exposure category.[5] The ILO emphasizes that transformation is generally more plausible than full job redundancy because many occupations combine tasks with different automation potential and still require human input. Exposure should therefore be read as a map of where work content may change, not as a forecast of how many jobs will disappear.

# 4. A practical framework for responsible use

## Figure 2. The human-AI work loop




*Five stages for converting a promising AI use case into a controlled business workflow.*

Source: Original BizNewsFeed framework informed by NIST AI RMF and NIST GenAI Profile [6,8].

**Takeaway: Governance is strongest when verification and ownership are designed into the workflow before scale.**

# 5. Decide by task, not by enthusiasm

| Task characteristic | Lower-risk / stronger fit | Higher-risk / weaker fit | Control to add |
|---|---|---|---|
| Ground truth | Inputs and correct answers can be checked | Correct answer is uncertain or inaccessible | Independent source checks and expert review |
| Cost of error | Low and reversible | Financial, safety, legal or reputational harm | Human approval, escalation and audit trail |
| Data sensitivity | Public or approved internal data | Personal, confidential or regulated data | Data classification, vendor controls, minimization |
| Novelty | Patterned or repetitive work | Rare edge cases or novel strategy | Pilot, red-team cases and fallback process |
| Output use | Draft or brainstorming input | Final decision or external factual claim | Named accountable owner and release checklist |

**Table 1. Task-suitability decision matrix.** Source: original synthesis based on NIST risk-management concepts [6,8] and evidence on heterogeneous task performance [2-4].

## 6. Measurement: prove value locally

A useful AI business case compares a defined baseline with an AI-assisted workflow. Measure time, throughput, quality, rework, error severity and user experience before and after introduction. For high-impact activities, track both average performance and tail risk: a small average gain can be unattractive if rare failures become materially worse. Avoid measuring only “hours saved,” because saved time has economic value only if it is redeployed productively or improves service, speed, quality or capacity.

Pilots should also distinguish tool capability from implementation quality. Weak prompts, poor source access, unclear ownership and missing review can make a capable model look ineffective; conversely, a polished demo can conceal failures that surface only at scale. NIST’s GenAI Profile identifies risks including confabulation, privacy, information security, intellectual property, harmful bias, overreliance and misuse, and recommends governance, measurement and management controls rather than reliance on disclaimers alone.[6]

## 7. Procurement and governance questions

When a generative-AI tool moves from individual experimentation to an enterprise workflow, procurement questions become part of the evidence problem. Teams should document which data may be submitted, whether prompts or outputs are retained, whether customer data may be used for model improvement, what security controls and subprocessors apply, and how model or product changes are communicated. A strong contract cannot prove that a model is accurate, but it can reduce avoidable uncertainty about data handling, service continuity and accountability. NIST's risk-management approach is useful because it treats governance as an organizational function rather than a feature inside the model.[6,8]

Organizations should also define a change-control threshold. If a provider replaces the underlying model, adds web access, changes retrieval sources or alters safety settings, prior evaluation results may no longer describe the deployed system. For material workflows, re-test after meaningful changes. Finally, keep an exit path: preserve source data, process documentation and a non-AI fallback for critical tasks. Vendor dependence is a business risk even when the AI performs well.

## 8. Common misconceptions

- “A model that sounds confident is probably correct.” Fluency is not a reliability metric; factual claims still require verification.
- “Productivity gains from one study apply to every job.” Experimental effects are task-, worker-, model- and workflow-specific.
- “AI exposure means a job will be automated.” Exposure describes task overlap with model capabilities; organizational choices, demand, regulation and complementary human work affect outcomes.[5]
- “Human in the loop” is enough. Review must be specific: who checks what, against which source, with what escalation rule and accountability?

## Reader checklist: an AI use-case gate

☐ Write the task and success metric in one sentence.

☐ Identify the authoritative source of truth for verification.

☐ Classify data that will enter the model and confirm it is permitted.

☐ Estimate the consequence of a wrong answer, not only its frequency.

☐ Test representative and adversarial cases before rollout.

☐ Assign a named human owner for final decisions and external claims.

☐ Measure quality and rework alongside speed.

☐ Document incidents, user feedback and material model or workflow changes.

## Conclusion

The evidence supports neither AI boosterism nor blanket skepticism. Generative AI has produced meaningful productivity and quality gains in controlled studies, especially on bounded language-intensive tasks and for less-experienced workers. It has also produced worse results when users relied on it outside a tested capability frontier. The business opportunity is therefore managerial: select tasks carefully, prove value with local measurement, and design verification, accountability and monitoring into the workflow. That approach treats AI as a capability to govern rather than a promise to believe.

## Research method and limitations

This paper is a narrative synthesis of peer-reviewed research, working papers, official statistics and risk-management guidance available as of 12 September 2026. It prioritizes randomized or field evidence where available. Limitations include rapid model change, differences in study tasks and populations, and the fact that firm-level adoption data use national survey definitions that are not perfectly uniform. Reported study effects should not be extrapolated to other tasks, industries or organizations without local testing.

## References

1. OECD, “AI use by individuals surges across the OECD as adoption by firms continues to expand,” 28 Jan 2026. Source link

2. Noy, S. & Zhang, W., “Experimental evidence on the productivity effects of generative artificial intelligence,” Science 381 (2023). Source link

3. Brynjolfsson, E., Li, D. & Raymond, L., “Generative AI at Work,” NBER Working Paper 31161. Source link

4. Harvard Business School / BCG field experiment, “Navigating the Jagged Technological Frontier” (2023). Source link

5. International Labour Organization, “Generative AI and jobs: A 2025 update”. Source link

6. NIST, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1. Source link

7. OECD/BCG/INSEAD, The Adoption of Artificial Intelligence in Firms (2025). Source link

8. NIST, AI Risk Management Framework 1.0 and implementation resources. Source link

# Appendix A. Underlying chart data

| Year | Share of firms using AI | Geography / coverage |
|---|---|---|
| 2023 | 8.7% | OECD countries where data available |
| 2024 | 14.2% | OECD countries where data available |
| 2025 | 20.2% | OECD countries where data available |

https://biznewsfeed.com