

THE AI VALUE GAP

From adoption to measurable business results

A Finland evidence base with European, United States and selected global evidence on the pathway from AI adoption to measurable business results.

Research cut-off: 9 September 2026

MandateMark™ Evidence & Insights publication.

Executive summary

Artificial intelligence can create real value. The evidence is increasingly clear on that point. Randomised field experiments show gains in professional work, software development and online retail; firm-level studies also find productivity and financial effects in particular settings. The harder organisational question is what happens between adopting AI and being able to establish a business result.

This report examines that distance. It does not treat AI access, adoption, integration, task productivity, operational improvement and business value as interchangeable. A worker completing a task faster is evidence of a task-level gain. It is not automatically evidence of lower organisational cost, higher revenue, better customer outcomes or stronger profitability. Those later claims require their own evidence.

The evidence also shows substantial heterogeneity. Effects differ by task, workflow, sector, firm size, worker experience and the complementary investments surrounding the technology. In one large online-retail platform, seven AI-enabled workflows produced effects ranging from no detectable sales effect to a 16.3% increase. In a field experiment with consultants, AI improved speed, completion and quality on tasks inside its capability frontier, but reduced correctness on a task outside it. [1][2]

The resulting MandateMark™ proposition is therefore not that organisations have failed to obtain value from AI. It is that a value pathway can narrow at several points: adoption may remain shallow; integration may not change work; task gains may not become operational gains; operational gains may not become business results; and genuine results may remain difficult to establish because baselines, costs, time horizons or attribution are weak.

Six findings

- AI can create measurable value, but the magnitude and type of value vary substantially by use case and organisational setting.
- Adoption is a poor endpoint. Firm use, business-function deployment and worker-task use are different measures.
- Task productivity is not the same as organisational value. Time saved or output increased must be traced to an intended operational or business result.

- Task fit and quality belong inside the value equation. Faster output can coexist with higher error risk when AI is used beyond its capability frontier.
- Complementary capabilities — including data, software, skills, process design and organisational capacity — influence whether adoption becomes productive.
- Measurement failure and value failure are different findings. A result may be real but poorly measured; equally, weak measurement cannot be used as proof that value exists.

1. What this report examines

The central research question is: As organisations invest in and adopt artificial intelligence, what determines whether that adoption translates into measurable business value — and how can leaders establish what actually changed?

The report uses Finland as its primary evidence context, with European and United States comparative evidence and selected global studies where they improve the explanation. It is not an AI-tools guide, a technology forecast, an AI Act compliance report or an argument that every organisation should adopt more AI.

A value pathway, not a binary verdict

The evidence supports examining one pathway with several possible breakpoints:

INTENT → ADOPTION → INTEGRATION → WORK REDESIGN → OPERATIONAL RESULT → BUSINESS RESULT → EVIDENCE → DECISION

The purpose is not to assign a universal maturity score. It is to identify where the pathway narrows for a particular organisation, commitment or use case, and what evidence is required before the next decision is justified.

2. Adoption is accelerating — but adoption is not integration

AI use is spreading rapidly, but headline adoption rates conceal large differences in firm size, sector, business function and intensity. Eurostat reports that 20% of EU enterprises used AI technologies in 2025, with substantially higher use among large enterprises than SMEs. Finland was among Europe's highest-adoption countries. [3]

The Finnish SME Barometer 1/2026 reports that 36% of responding SMEs use some generative AI and that lack of knowledge or expertise is the most frequently reported obstacle. This is important evidence of diffusion and capability constraints; it does not establish what value those firms obtain from use. [4]

United States Census research makes the measurement problem especially clear. It distinguishes whether a firm uses AI at all, how many business functions deploy it, and how extensively workers use it in tasks. In the November 2025–January 2026 reference period, 18% of firms reported AI use in a business function, while employment-weighted use was 32%. Many adopters used AI in only a limited number of functions. [5]

The implication

An organisation can truthfully be an AI adopter while remaining lightly integrated. “We use AI” therefore says very little about whether AI has changed the operating model or produced a measurable result.

3. AI can improve performance — sometimes substantially

The evidence base now contains multiple studies in which AI produces meaningful measured gains. Three randomised field experiments involving software developers at Microsoft, Accenture and a Fortune 100 company found an aggregate 26.08% increase in completed tasks among developers given access to an AI coding assistant, with larger gains among less-experienced developers. [6]

A large online-retail study integrated generative AI into seven consumer-facing workflows and randomised exposure across millions of users and products. Effects on sales ranged from no detectable effect in some workflows to 16.3% in the strongest application. Because prices and workflow inputs were held constant across treatment and control conditions, the authors interpret the positive effects as productivity improvements. [1]

In asset management, a 2026 Review of Financial Studies article finds that hedge funds adopting generative AI earned 2–4% higher annualised abnormal returns than non-adopters in its difference-in-differences design, while non-hedge funds did not show the same benefit. AI talent and the type of information work performed appear relevant to the result. [7]

These studies are important because they prevent an easy but inaccurate narrative that AI investment is generally failing. The evidence instead points to conditional value: AI can produce gains, but where, how and for whom matters.

4. Task value is not business value

A recurring weakness in AI-value discussions is the movement from a task metric directly to a business claim. If a worker drafts a document 30% faster, the organisation has evidence of a task-level change. It does not yet know whether the saved time became additional productive capacity, reduced labour cost, improved customer service, faster delivery or no material organisational change.

A randomised field experiment covering 7,137 knowledge workers across 66 firms found meaningful individual time savings from generative AI, but did not detect broader changes in the quantity or composition of work during the study period. The finding does not mean the saved time had no value; it shows why aggregation from individual productivity to organisational result cannot be assumed. [8]

Four levels of value

TASK VALUE Did AI improve speed, quality or output on a defined task?

OPERATIONAL VALUE Did those gains change a workflow, process, capacity, cost or service result?

BUSINESS VALUE Did the operational change contribute to revenue, margin, growth, innovation, customer outcomes or another intended organisational result?

EVIDENTIARY VALUE Can the organisation establish what changed, against what baseline, at what cost, over what period and with what confidence?

5. The capability frontier: more output can also mean more error

Value cannot be assessed through speed and volume alone. In a field experiment with 758 consultants, participants using GPT-4 completed 12.2% more tasks, worked 25.1% faster and produced higher-quality results on tasks within the model's capability frontier. On a task deliberately selected outside that frontier, however, AI users were 19 percentage points less likely to reach the correct answer. [2]

This finding has a direct organisational implication: the value equation must include task fit, quality, verification and the consequences of error. An intervention that increases throughput but also increases costly mistakes may create less net value than the productivity metric suggests.

Human oversight therefore should not be treated only as a governance cost. In some applications it is part of the production system required to convert AI output into reliable organisational output.

6. Integration and complementary capability

The evidence increasingly suggests that technology access alone is insufficient. EIB research using matched firm-level data from more than 12,000 non-financial firms in Europe and the United States estimates that AI adoption increased labour productivity by roughly 4% in European firms. Gains were concentrated more strongly among medium and large firms, and complementary investments in software, data and workforce training were important. [9]

OECD research similarly identifies skills, finance, data, connectivity and implementation capacity as important conditions for SME adoption and use. Its 2026 SME evidence finds rapid uptake alongside uneven strategic and secure integration. [10]

These findings do not prove that every organisation must undertake a large transformation programme before AI can create value. Some narrow use cases generate strong results. They do show why procurement alone is an inadequate theory of value creation.

7. Size and sector change the pathway

AI adoption and value are heterogeneous. Larger firms generally report higher adoption, but scale can bring both advantages and complexity: more data and capital, but also more systems, governance requirements and organisational dependencies. SMEs may have fewer resources for integration while also being able to change a narrow workflow more quickly.

Sector differences are equally important. AI's measurable contribution depends on the nature of work, data availability, error costs, customer interaction, regulation and whether the relevant task is already highly digitised. The asset-management evidence, software-development experiments and online-retail experiments demonstrate very different mechanisms of value. [1][6][7]

MandateMark therefore should not compare an SME, municipality, retailer and multinational financial institution against one universal AI-value benchmark. The relevant baseline is the organisation's intended result and the specific workflow or capability being changed.

8. The cost side of value

The price of an AI licence is not the total cost of creating an AI-enabled result. Depending on the use case, the relevant cost base can include data preparation, integration, infrastructure, evaluation, security, governance, workforce training, workflow redesign, human verification, maintenance and change management.

OECD productivity work notes that complementary intangible investments associated with technology adoption can be difficult to capture fully in conventional accounts. EIB evidence likewise points to software, data and training as complements to productive AI adoption. [9][11]

This means an organisation should resist calculating return from a narrow technology denominator when significant implementation costs sit elsewhere in the business. Equally, it should not inflate costs by assigning every adjacent digital investment to AI. The appropriate boundary must be explicit and consistent.

A practical accounting question

What did the organisation spend in order to create the result — not merely what did the AI tool cost?

9. Time-to-value and the danger of premature verdicts

AI investments can produce immediate task effects while broader organisational results emerge later. Integration, training, workflow redesign and behavioural adaptation can create a lag between expenditure and measurable outcome. This creates two opposite risks: declaring success from an early activity metric, or declaring failure before the relevant result could reasonably appear.

BEA research on AI expectations and outcomes finds that the relationship between firms' stated motivations and observed outcomes remains difficult to interpret and that structural change may precede visible outcome measures. [12]

MandateMark therefore distinguishes NOT ACHIEVED from NOT YET DETERMINABLE. A time horizon should be defined when the commitment is made, not retrofitted after disappointing results.

10. Measurement: reported, demonstrated and expected value

AI-value evidence frequently mixes three different categories. They should be separated.

DEMONSTRATED RESULT A measured change supported by an experimental, quasi-experimental, administrative, financial or otherwise appropriate design.

REPORTED RESULT A benefit reported by executives, workers, customers or the organisation itself.

EXPECTED RESULT A projected or anticipated future benefit.

All three can be useful. They do not carry the same evidential weight. A CEO survey reporting positive ROI is evidence of executive assessment; it is not equivalent to audited financial return or a controlled estimate of causal impact.

This distinction becomes particularly important because AI-adopting firms may differ systematically from non-adopters before AI is introduced. More productive, better-capitalised or more digitally mature organisations may also be more likely to adopt. Simple comparisons can therefore overstate what can be attributed to AI.

11. The AI value pathway

The evidence supports a disciplined leadership sequence rather than a universal ROI score.

INTENT What result is the organisation trying to produce, and why is AI relevant to it?

ADOPTION Who is actually using AI, for what work and with what frequency?

INTEGRATION In which business functions and workflows is AI embedded?

WORK REDESIGN What changed in the way work is allocated, performed, checked or handed off?

OPERATIONAL RESULT What changed in speed, quality, output, capacity, cost, service or another operational measure?

BUSINESS RESULT Did the operational change contribute to the intended organisational result?

EVIDENCE What baseline, comparison, cost boundary and attribution method support the claim?

DECISION What should happen next: SCALE, ADAPT, REDESIGN, CONTINUE, STOP or INVESTIGATE?

12. What leaders should ask

- What business result did we intend AI to change?
- Where is AI actually being used, and where is use merely nominal?
- What changed in the workflow because AI was introduced?
- What task-level and operational effects can we observe?
- Did those effects reach a business result?
- What did the full intervention cost, including complementary investment and control?
- What quality, risk or workforce trade-offs accompanied the gain?
- What is demonstrated, what is reported, what is expected and what remains unestablished?
- What decision does the evidence justify now?

13. Implications for Finland

Finland is a particularly useful evidence context because enterprise AI adoption is comparatively high while the SME evidence still shows capability constraints. High national adoption should not be confused with uniform organisational integration or value capture. The next competitive question is therefore less about whether Finnish organisations encounter AI and more about whether they can select appropriate uses, integrate them into work, measure the resulting changes and make disciplined follow-on decisions. [3][4]

For SMEs, proportionality is essential. The measurement system should be strong enough to support a decision without becoming an administrative burden larger than the use case itself. A small firm may need a clear baseline, a bounded workflow, a few operational and business measures, a cost boundary and a review date — not an enterprise transformation office.

14. Conclusion: from adoption to evidence

The evidence reviewed for this report does not support a simple verdict on AI. It supports something more useful: AI can create substantial value, but value is conditional, uneven and frequently difficult to trace from use to organisational result.

The central challenge is therefore not proving that AI works in the abstract. It is establishing whether a particular organisation, applying AI to particular work, produced the result it intended — at an acceptable cost and risk — and whether the evidence is strong enough to justify the next decision.

That is the AI value gap: the distance that can exist between adoption, AI-enabled change, operational result, business result and evidence.

For leaders, the practical question is simple:

You use AI. What changed?

Can you establish it?

Evidence note

This publication distinguishes official statistics, academic and experimental research, institutional comparative evidence and stakeholder or consultancy material. Headline claims are anchored primarily in official, peer-reviewed, experimental or high-quality institutional sources. Association is not presented as causation where the underlying research design does not support it. Percentages from surveys with different populations, definitions and reference periods are not treated as directly comparable.

Selected references

- [1] Fang, L., Yuan, Z., Zhang, K., Donati, D. & Sarvary, M. Generative AI and Firm Productivity: Field Experiments in Online Retail. Columbia Business School, forthcoming working paper.
- [2] Dell’Acqua, F. et al. The Cybernetic Teammate / Navigating the Jagged Technological Frontier evidence, field experiment with consultants; Organization Science, 2026 publication context.
- [3] Eurostat. Digitalisation in Europe — 2026 interactive publication; enterprise use of artificial intelligence, 2025 data.
- [4] Federation of Finnish Enterprises et al. SME Barometer 1/2026.
- [5] U.S. Census Bureau. The Microstructure of AI Diffusion: Firm Adoption, Worker Use, and Business Function Deployment. CES Working Paper, 2026.
- [6] Cui, Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S. & Salz, T. The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. Management Science, 2026.
- [7] Generative AI and Asset Management. The Review of Financial Studies, published May 2026.
- [8] NBER working paper, 2026: randomised experiment with 7,137 knowledge workers across 66 firms examining generative-AI use and time allocation.
- [9] European Investment Bank / BIS. AI Adoption, Productivity and Employment: Evidence from European Firms. 2026.

[10] OECD. Empowering SMEs in the Age of AI; AI Adoption by Small and Medium-Sized Enterprises. 2026.

[11] OECD. Compendium of Productivity Indicators 2026.

[12] U.S. Bureau of Economic Analysis. AI Expectations and Outcomes. 2026.

[13] Eisfeldt, A. L., Schubert, G., Taska, B. & Zhang, M. B. Generative AI and Firm Values. Journal of Finance, forthcoming 2026 / NBER Working Paper 31222.

[14] Stanford HAI. AI Index Report 2026 — Economy chapter.

About MandateMark™

MandateMark™ is an evidence-led methodology for organisations that need to move a defined commitment from intention into accountable delivery and measurable results. Its work connects evidence, diagnosis, decision, implementation and review while distinguishing what the evidence establishes from what remains uncertain.

Evidence & Insights publications are public research resources. Detailed MandateMark™ diagnostic instruments, facilitator methods and client-specific materials are proprietary and are not publicly distributed.