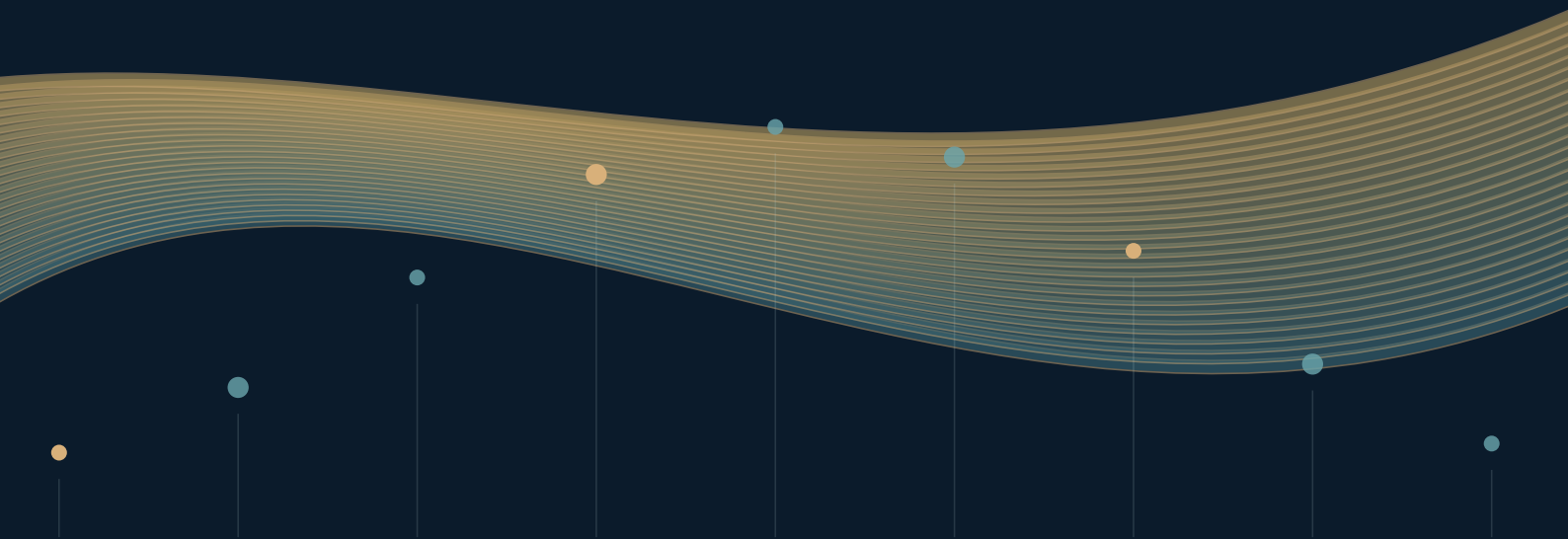




CROSS-SECTOR RESEARCH

What makes an AI-native enterprise?

A research synthesis on work, organizational capability and the evidence needed to measure them.



ORIGINAL MAYDEN GRAPHIC / EVIDENCE IN MOTION

01 / THE EXECUTIVE VIEW

Start with the operating evidence.

A research synthesis on work, organizational capability and the evidence needed to measure them.

80

selected entity records

272

recorded claims

224

eligible claims

21

unranked records

01 / WHAT THE RECORD SHOWS

59 of 80 selected entity records contain qualifying AI operating evidence in this review. 21 remain unranked because no qualifying claim was located; this does not establish absent capability.

02 / WHAT THE RECORD SHOWS

The corpus contains 272 recorded claims, of which 224 meet the declared eligibility rules. Specific commitments, deployed practices and operating detail are coded separately.

03 / WHAT THE RECORD SHOWS

Decision Lineage appears in 4 of these 80 profiles, compared with 44 for AI Core. These are counts within the selected research corpus, not enterprise adoption rates.

Public-evidence review. Selected entity scopes. No company participation, population inference or independent assurance implied.

Read the argument, inspect the exhibits, then use the leadership questions to identify the next evidence your organisation should produce. All report bodies are in English; Arabic executive summaries are available online.

EXHIBIT 01 / PUBLIC EVIDENCE

Where evidence is visible across the eight sector cohorts

Number of selected entity records with a qualifying claim in each theme; cohort size shown per row

	Core	People	Partners	Interface	Lineage	Testing	Authority	Knowledge	Control	
Aviation & Logistics	7	3	5	3	1	2	0	2	2	/ 10
Banking	7	8	4	7	2	3	2	4	7	/ 10
Energy & Utilities	7	6	5	3	0	3	1	3	1	/ 10
Healthcare	6	3	3	4	0	3	0	1	4	/ 10
Manufacturing & Materials	5	2	2	1	0	3	0	2	2	/ 10
Real Estate & Construction	4	1	4	3	0	3	1	1	0	/ 10
Retail & Consumer	3	1	3	5	0	0	1	1	2	/ 10
Telecommunications	5	5	3	4	1	2	3	3	3	/ 10

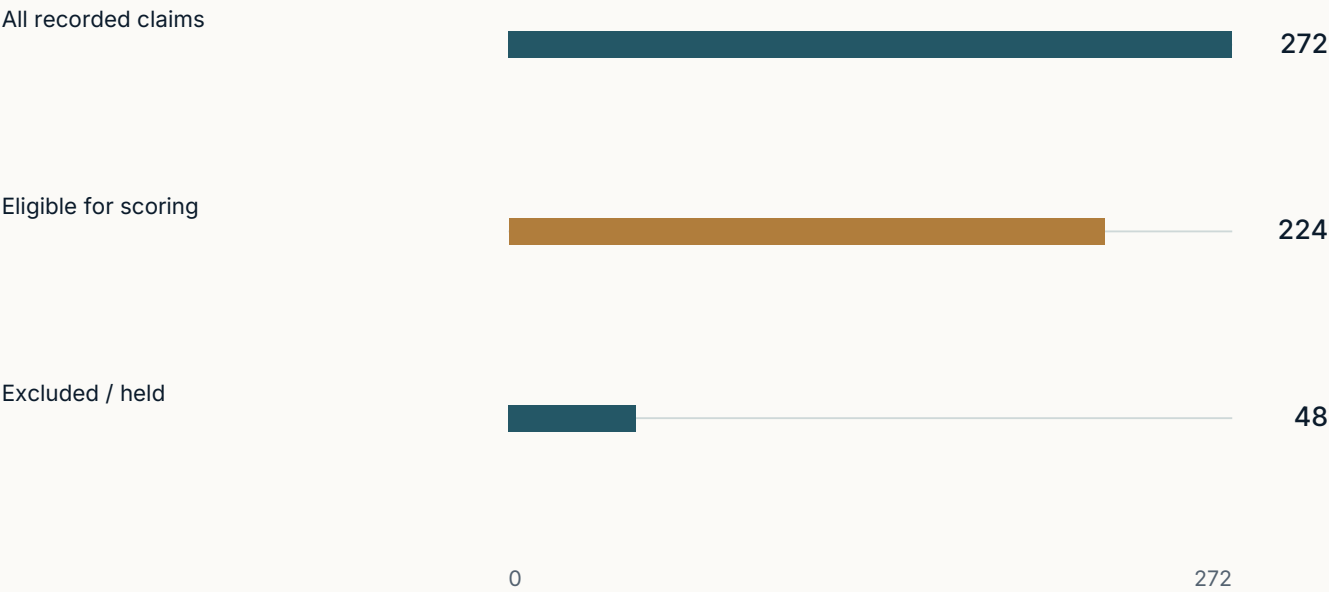
Cells show counts within each named cohort; / n is the cohort denominator.

Source: frozen Mayden public-evidence corpus. Counts describe located evidence, not comparable sector maturity or prevalence. Cohorts are purposeful, heterogeneous and sometimes overlap.

EXHIBIT 02 / PUBLIC EVIDENCE

The evidence record includes what did not qualify

Recorded claims pass explicit access, scope, deployment and review gates



Source: release claim ledger. These counts do not measure research recall, factual accuracy or company capability. All exclusions remain inspectable in the public dataset.

02 / ANALYSIS

The operating model matters more than the announcement

Mayden AI Native Index · Research foundation · 10 October 2026

An AI-native enterprise repeatedly converts AI capability into useful, accountable work across a defined organizational scope. This is Mayden's proposed working definition, to be tested through field research. It asks what changes in decisions, workflows, responsibilities and learning when AI becomes part of normal operations.

A company can have thousands of AI licences and little operating change. Another can use a small number of models inside essential workflows, with clear ownership, sound evidence and effective fallback arrangements. A benchmark that rewards purchasing activity will miss this difference. A benchmark that rewards autonomy without considering consequence may reward poor judgment.

The research offers a strong foundation for measuring these distinctions. It does not yet provide a universally validated definition of an AI-native enterprise. Mayden's opportunity is to build that measurement carefully, beginning with Saudi Arabia, the UAE, Qatar, Kuwait and Bahrain.

Productivity gains are real, conditional and uneven

The published customer-support study by Brynjolfsson, Li and Raymond follows 5,172 agents. AI assistance increased resolved issues per hour by 15% on average, with larger gains among less experienced workers. This is evidence about a particular deployed workflow; it is not a forecast for every company's profit margin. The published figures differ from earlier working-paper versions, a reminder that a research database needs version control. [Published study](#).

In a randomized experiment with 453 professionals, Noy and Zhang found faster completion and higher assessed quality on writing tasks. That

demonstrates potential at the task level. It does not incorporate every approval, integration, legal review or downstream correction in an enterprise process. The Index should therefore ask whether faster drafting improved the accepted final output and the total process. [Writing experiment](#).

The consulting experiment involving 758 BCG professionals reinforces the need to evaluate task suitability. Its findings describe a jagged boundary: apparently similar tasks can sit on different sides of what the tool handles reliably. A serious company must know where its deployed systems succeed and where users need another approach. Counting deployments cannot reveal that knowledge. [HBS working paper](#).

The executive question is precise: **Which material workflow performs better, against which baseline, after counting the work needed to make its outputs usable?**

AI can change expertise and collaboration

In the Procter & Gamble experiment, 776 professionals worked on product innovation challenges. Individuals equipped with AI matched the assessed performance of teams without AI. The study opens an important question about how expertise can travel across functional boundaries. It does not establish that companies should remove teams or that temporary experimental performance will persist in ordinary operations. [Cybernetic Teammate study](#).

A separate six-month experiment with 7,137 knowledge workers found changes in email and document work without a significant change in meeting time in its May 2025 revision. The contrast points toward a useful hypothesis: individual changes can be easier than coordinated organizational changes. Giving people tools does not itself settle decision rights, redesign handoffs or change shared routines. [Pinned working-paper revision](#).

Our next company study will test that hypothesis through actual operating examples. Who now prepares a recommendation? Who verifies it? Who can act on it? Which handoff has disappeared, and which new review step has appeared? A before-and-after workflow can reveal more than a general claim that employees are empowered.

Perceived improvement is not measured improvement

METR's early-2025 experiment found that 16 experienced open-source developers took longer on AI-allowed tasks, despite believing that the tools accelerated them. The result is bounded by its selected developers, mature repositories and tools of that period. It is useful evidence against treating users' perceived time savings as an objective productivity measure. [Original experiment](#).

The same research organization later explained why its follow-up estimates were unreliable. Developers' willingness to participate, their task selection and time accounting changed. Its February 2026 report therefore cannot support a clean claim that the original effect has reversed by a precise amount. This is a lesson for any longitudinal index: the people and work entering the measurement may change alongside the technology. [Follow-up methods report](#).

A 2026 survey of almost 6,000 executives across four non-GCC economies also separates widespread reported AI use from limited reported realized effects. Executive expectations were more optimistic than descriptions of past results. These are survey reports, not audited financial effects. For Mayden, the implication is to record use, measured outcomes and future expectations in separate fields. [February 2026 author version](#).

Better combinations must be demonstrated

An existing meta-analysis of 106 experiments found that human-AI combinations, on average in that corpus, underperformed the better of the human or AI working alone. Results differed by task type. This is not the claim that AI assistance never helps

people; the comparison is with the better standalone performer. Its study window ended in June 2023, so it cannot establish the performance of today's systems. [Published meta-analysis](#).

The practical lesson is to test the combination. Human review needs enough information, time and authority to improve the result. A nominal approval click should not be treated as effective oversight. Equally, the correct design may deliberately retain human judgment even where automation is faster, because accountability and consequence matter.

Mayden's score should reward an appropriate operating design. It should not award automatic bonus points for replacing people, maximizing autonomy or adopting a particular model.

Capability includes organizational resources

Mikalef and Gupta developed a multidimensional AI capability instrument covering technical, human and organizational resources. Their survey provides useful construct-development precedent, but its associations with performance are not causal proof and its sample predates the current generative-AI environment. This work informs our question design; local validation remains necessary. [AI capability study](#).

The productivity J-curve literature explains why complementary investments in processes and skills can complicate the timing of measured gains. It provides a plausible mechanism, not an excuse for an unproductive project to continue indefinitely. Assessment should record implementation costs, expected timing, milestones and explicit stopping criteria. [J-curve research](#).

Organizational measurement also has an established precedent in the management-practices research of Bloom and Van Reenen. Structured evidence about how firms operate can be related to performance. That supports asking for operational examples and training assessors. It does not make a newly authored AI score valid by association. [Management measurement study](#).

Nine questions that executives should be able to answer

Mayden's existing nine themes can form a coherent research agenda. They remain candidate

dimensions, with overlap and weighting to be tested.

Exhibit 03 - Nine operating questions for the enterprise

Theme	The executive question
AI Core	Where does AI support material recurring work, using reliable data and a workable fallback?
Workforce	Can people use, challenge and improve AI in the roles they actually perform?
Ecosystem	Can we evaluate, maintain and, when necessary, change critical AI dependencies?
Interfaces	Is AI part of the workflow, and what share of eligible work actually uses it?
Decision lineage	Can we reconstruct an important output, its sources and the actions that followed?
Experimentation	What evidence causes us to scale, revise or stop a deployment?
Operating model	Who owns the outcome and can authorize, halt or reverse consequential actions?
Knowledge	Who keeps the organization's usable knowledge current, accessible and trustworthy?
Governance	Can we discover uses, evaluate risks and respond to incidents across the lifecycle?

These questions are deliberately about behavior. A policy can contribute evidence, but it does not prove implementation. A training course can contribute evidence, but it does not prove competence. A partnership can provide resources, but it does not prove a deployed operating practice.

NIST's risk-management framework offers useful structure for lifecycle controls and responsibilities. It

is guidance, not a certificate of compliance or a validated productivity score. **AI RMF 1.0**.

A credible benchmark separates three records

First is **operating capability**: what practices exist, over what scope and with what continuity. Second is **outcome evidence**: changes in quality, speed,

cost, revenue, risk and work experience. Third is **evidence strength**: the basis for believing each statement.

Keeping these records separate prevents a circular argument in which high-performing companies receive high capability scores because their outcomes were included in the score, and the same outcomes are then used to claim the score predicts performance.

It also prevents public visibility from masquerading as capability. A company with a detailed annual report may be easier to assess than a private firm. Lack of published evidence is an information gap. It is not proof that a practice is absent. The initial public-document product must therefore identify itself as a **Public Evidence Index**, with explicit limits on what its ranking means.

Normalization cannot repair a weak construct. The OECD/JRC handbook makes weighting, aggregation and sensitivity part of the design problem. A fixed scale is useful for repeat measurement, but the anchors still need justification. **Composite-indicator handbook**.

Build a regional benchmark that can travel

The five-market starting scope is a deliberate launch cohort, not a claim to cover the entire GCC. Country, sector, company size, ownership and operating scope should be visible. A regional subsidiary must not inherit a global parent's capabilities without evidence that those practices apply locally.

Arabic and English instruments require more than literal translation. Participants should explain what they understood each question to mean; researchers should investigate differences by language and interview mode. Voice answers need confirmation before being transformed into scores. Scale-development guidance supports an explicit progression from item design to testing and validation. **Measurement primer**.

The valuable longitudinal question is not simply who ranks first this year. It is **which operating changes predict sustained improvements, under which conditions, and which apparently promising practices fail to deliver**. Answering it requires repeat participants, preserved instrument versions, comparable outcomes and honest treatment of firms that stop responding.

A useful confidential report can help sustain that participation. It should show an organization's supported strengths, unresolved questions and a few concrete priorities. The public research can then explain patterns without exposing entrusted company evidence.

Research note

This report is a scoped rapid evidence review and Mayden design synthesis. It is not a systematic review or a new statistical meta-analysis. Fifteen research and methodological records were selected through targeted primary-source searches on 10 October 2026. Selection prioritized identifiable original studies, peer-reviewed measurement work and authoritative methodological guidance; it was not an exhaustive database search.

Experiments, observational studies, executive surveys, an existing meta-analysis and frameworks are kept distinct. No pooled productivity percentage is calculated. Source access ranged from publisher/author abstracts and indexed passages to relevant report sections; this is not a claim that every full paper and appendix was independently audited. Publication versions are pinned where findings changed. None of these sources establishes GCC company rankings, validates Mayden's nine dimensions or replaces proprietary fieldwork.

Review transparency. Extraction and source challenge in this edition were AI-assisted. Separate AI readers do not constitute human expert sign-off or independent assurance. Company claims, research findings and Mayden interpretations retain their distinct scope.

EXHIBIT 04 / INSPECT THE ASSUMPTION

Operating detail deserves a separate test

Eligible claim records, n = 224; disclosure anchors are not maturity stages



Source: reviewed coding. Detail is not independent assurance, and multiple claim records may describe a shared underlying practice.

EXHIBIT 05 / INSPECT THE ASSUMPTION

Build a benchmark from evidence to a decision

Mayden research architecture; conceptual framework, not an observed causal model

- 01** Define Entity, workflow, population and period.
- 02** Observe Sources, tasks, usage and operating boundaries.
- 03** Challenge Alternative explanations, controls and exclusions.
- 04** Calculate Declared rules, sensitivity and reproducibility.
- 05** Decide Specific action, evidence gap and next test.

Source: Mayden methodological synthesis. Confidential company research remains a separate instrument requiring permission, respondent testing and validation.

04 / THE LEADERSHIP AGENDA

Ask for evidence that changes the decision.

01 / Choose one recurring workflow.

Evidence to request: Its users, decisions, frequency, business importance and operating boundary.

02 / Make authority and evidence inspectable.

Evidence to request: Input provenance, evaluations, interventions, accountable owner and recovery path.

03 / Track value with a credible comparator.

Evidence to request: Quality, cost, risk and adoption denominators over a stated period.

Method and boundaries

This is an AI-assisted, bounded public-source review using a purposeful corpus. Nine themes retain their highest eligible anchor: 1 for a specific commitment or pilot, 2 for an implemented practice, 3 for implemented practice with operating detail. U means unknown. Points equal 100 times the sum of anchors divided by 27. Ordinal steps and equal weights are design assumptions, not validated importance estimates.

Reported outcomes, actual company capability and disclosure depth are different objects. Source-linked corporate or vendor statements are not independently assured operations. Deterministic sensitivity tests are not confidence intervals. Sector scopes, reporting periods and access differ; no national rankings, population prevalence or causal inference are made. The confidential company instrument remains a separate development track.

Dated sources run from 2019-05-14 to 2026-10-01. 79 of 224 eligible claim records have undated sources. Publication dates do not confirm continuing operation or a common observation period; undated does not mean stale.

[Inspect the full method](#) / [Reproduce the evidence](#)

05 / SOURCES & PROVENANCE

Follow the argument back to its source.

Sources below include eligible and excluded research leads. Inclusion in this list does not make a source eligible or independently verified. Exact claim decisions, source locators and limitations are retained in the online evidence ledger. Research cutoff: 10 October 2026.

S01 / Generative AI at Work

Quarterly Journal of Economics / 2025-02-04 / peer-reviewed empirical

Locator: Staggered workplace rollout; observational causal design

[Open source](#)

S02 / Experimental evidence on the productivity effects of generative artificial intelligence

Science / 2023-07-13 / peer-reviewed experiment

Locator: Preregistered randomized online experiment

[Open source](#)

S03 / Navigating the Jagged Technological Frontier

Harvard Business School working paper 24-013 / 2023-09 / working-paper experiment

Locator: Randomized field experiment

[Open source](#)

S04 / The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise

NBER working paper 33641; published Organization Science June 2026 / 2025-04 / original field experiment

Locator: Preregistered randomized field experiment

[Open source](#)

S05 / Shifting Work Patterns with Generative AI

NBER working paper 33795, revision 1 / 2025-05-06 / working-paper experiment

Locator: Six-month randomized field experiment

[Open source](#)

S06 / Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity

METR / 2025-07-10 / original experiment

Locator: Task-randomized controlled trial

[Open source](#)

S07 / We are Changing our Developer Productivity Experiment Design

METR / 2026-02-24 / original follow-up methods report

Locator: Follow-up randomized study; methodological disclosure

[Open source](#)

S08 / Firm Data on AI

NBER working paper 34836; author-hosted February version / 2026-02 / original survey

Locator: Stratified firm surveys

[Open source](#)

S09 / When combinations of humans and AI are useful: A systematic review and meta-analysis

Nature Human Behaviour / 2024-10-28 / peer-reviewed evidence synthesis

Locator: Preregistered systematic review/meta-analysis

[Open source](#)

S10 / Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance

Information & Management 58(3) / 2021-04-01 / peer-reviewed survey and measurement

Locator: Instrument development and cross-sectional survey

[Open source](#)

S11 / The Productivity J-Curve: How Intangibles Complement General Purpose Technologies

NBER; American Economic Journal: Macroeconomics 13(1), 2021 / 2021-01 / peer-reviewed theory and empirical analysis

Locator: Economic model and historical empirical analysis

[Open source](#)

S12 / Measuring and Explaining Management Practices Across Firms and Countries

NBER; Quarterly Journal of Economics 122(4), 2007 / 2007-11
/ peer-reviewed organizational measurement

Locator: Structured management-practice interviews and
linked performance analysis

[Open source](#)

S13 / Artificial Intelligence Risk Management Framework (AI RMF 1.0)

NIST / 2023-01 / official framework

Locator: Voluntary risk-management framework

[Open source](#)

S14 / Handbook on Constructing Composite Indicators: Methodology and User Guide

OECD / European Union / EC-JRC / 2008-08-22 /
measurement guidance

Locator: Methodological handbook

[Open source](#)

S15 / Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer

Frontiers in Public Health / 2018-06-11 / peer-reviewed
measurement guidance

Locator: Methodological primer

[Open source](#)

About Mayden

Mayden connects senior judgement, research and engineering to help organisations move from strategic choices to working capabilities. This research programme examines how AI changes recurring work, decisions and accountability.

[Explore the Index](#) / [Submit a correction or discuss the evidence](#)