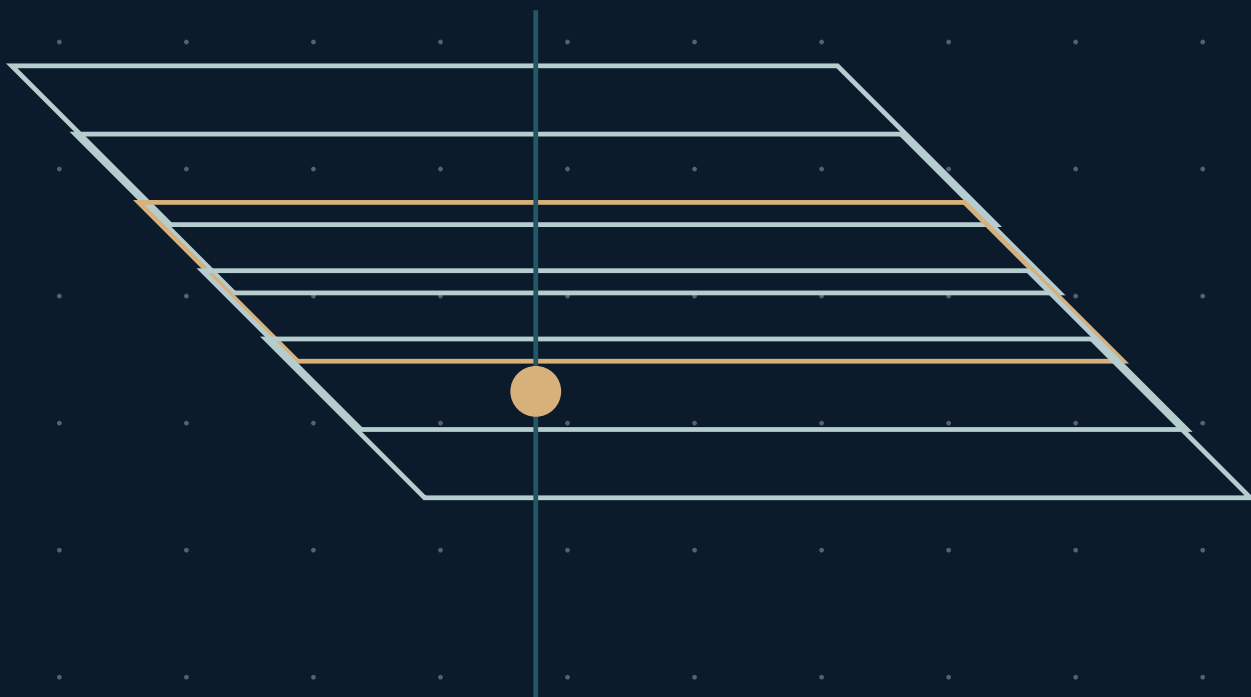


CROSS-SECTOR RESEARCH

How much should you trust an AI index?

Evidence quality, disclosure bias and what changes when scoring assumptions change.



ORIGINAL MAYDEN GRAPHIC / OPERATING QUESTIONS

01 / THE EXECUTIVE VIEW

An index earns trust through its construct, evidence and reproducibility, not its visual precision.

Evidence quality, disclosure bias and what changes when scoring assumptions change.

80

selected entity records

272

recorded claims

224

qualifying claim records

9

candidate research themes

01 / THE OPERATING QUESTION

An index needs an explicit construct, coherent units and evidence suited to its intended decision. Reproducible arithmetic addresses only one requirement.

02 / THE OPERATING QUESTION

Mayden now publishes sector research and attributed examples without individual company scores. Historical disclosure comparisons retain their original scope and limitations.

03 / THE OPERATING QUESTION

A future benchmark needs human content review, bilingual cognitive work, independent coding, empirical tests and governance that can stop publication. These gates remain incomplete.

AI-assisted, bounded public-source research. Purposefully selected entity scopes; no company participation, endorsement, independent assurance or population inference implied. Confidential assessment is in development.

02 / THE ARGUMENT

Four questions to carry into the business.

Read each step as a test of the operating claim. This is an analytical framework, not a measured maturity ladder.

- 1** **What does the number claim to measure?**
Normalisation makes observations calculable on a chosen scale. It does not establish that components represent the intended construct or deserve their weights.
- 2** **Which source dependencies remain?**
A document can support several claims; URLs can repeat one assertion; claims can describe one case; a group can appear through subsidiaries. URL volume is not independent confirmation.
- 3** **Which decisions can be replayed?**
The 11 October challenge tested 512 joint weight-corner specifications and 36 ordinal-spacing specifications against frozen v4, alongside a separately labelled baseline reference. Independent replay checked 43,840 specification-record results across 548 nonbaseline choices and 80 records.
- 4** **What validation must come next?**
Predeclare the intended use and test content, comprehension, coding, comparability and outcomes. Human appointments and empirical validation remain incomplete.

How to use this report

Follow the source accounts and their limits, use the diagrams to frame an evidence request, then choose one recurring workflow for deeper examination. A company name identifies an attributed account; it does not imply participation, endorsement or an enterprise rating.

03 / THE ANALYSIS

An index earns trust through its construct, evidence and reproducibility, not its visual precision.

Begin with the claim the number would make

Normalisation makes observations calculable on a chosen scale. It does not establish that components represent the intended construct or deserve their weights. The OECD/JRC handbook exposes such modelling choices; it does not validate Mayden's dimensions. [Handbook on Constructing Composite Indicators: Methodology and User Guide](#).

Historical public-evidence editions used the strongest qualifying disclosure in each of nine themes, a fixed denominator and equal nominal weights. Arithmetic could be replayed, but one narrow implementation could fill a theme, unknown public evidence contributed no documented points, and access varied. These properties limit enterprise-capability interpretation. The current edition removes individual company scores from active public research.

The future instrument should distinguish a formative composite, reflective subscales and direct descriptive measures. A formative design defines a concept through selected components; high correlation is not automatically desirable. Reflective components need their own model and tests. Neither is validated by a convenient distribution.

[Conventional Wisdom on Measurement: A Structural Equation Perspective; Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer](#).

Keep source, claim and operating case separate

A document can support several claims; URLs can repeat one assertion; claims can describe one case; a group can appear through subsidiaries. URL volume is not independent confirmation. Record original assertion, document version, extracted passage, claim, review and release. A provenance graph exposes derivation, not truth. [PROV-DM: The PROV Data Model](#).

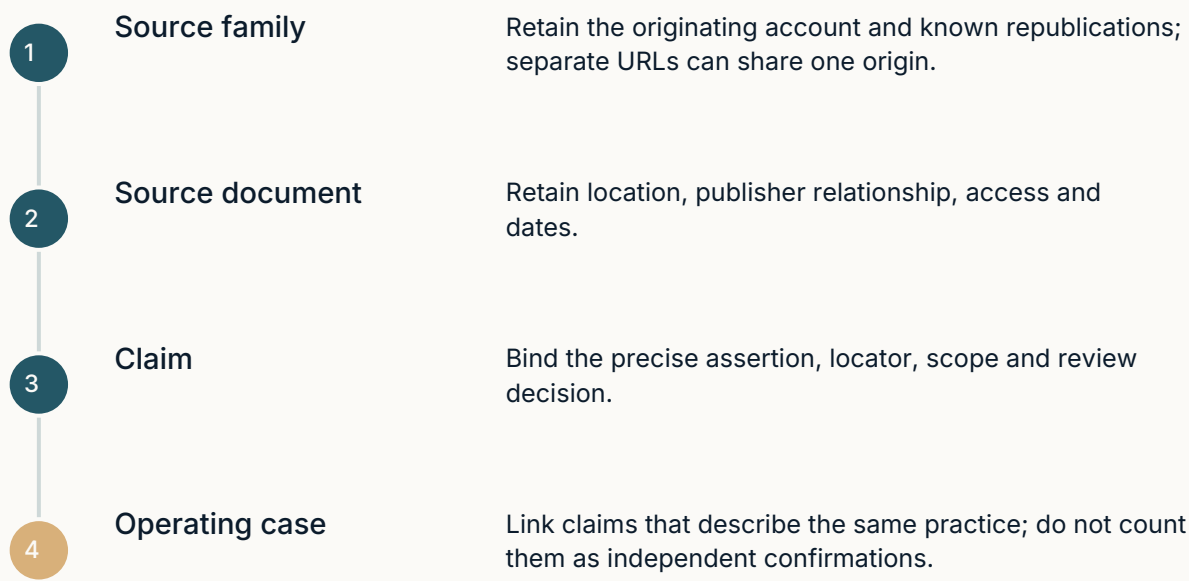
Legacy review records are not all bound to exact claim hashes, and some sources lack recorded dates. New controls should surface those gaps without rewriting history as if reviews had occurred. A queue cannot confer eligibility, confirm current operation or manufacture a second reviewer.

Missingness needs a reason: not searched, no qualifying evidence within a declared search, inaccessible source, outside scope or an observed negative finding. A participant instrument also needs unknown and not-applicable responses. Treating every state as zero creates precision about a different question.

EXHIBIT 01 / CONCEPTUAL TOOL

A source, a claim and an operating case are different units

Mayden analytical framework / questions to test, not measured maturity stages



Mayden conceptual framework. These questions specify evidence to request; they do not assert that a named organisation has or lacks the practice. This is not a validated instrument or an observed causal sequence.

Unit: Conceptual decision or evidence request. Research cutoff: 2026-10-10.

03 / THE ANALYSIS

Challenge modelling choices without inventing probabilities

The 11 October challenge tested 512 joint weight-corner specifications and 36 ordinal-spacing specifications against frozen v4, alongside a separately labelled baseline reference. Independent replay checked 43,840 specification-record results across 548 nonbaseline choices and 80 records. It verifies tested arithmetic, not source truth or construct validity.

Within the 512 weight choices, 53 of 59 historically ranked records changed position in at least one tested specification. Analysts selected the set; it is not a probability distribution, an exhaustive plausible model space or an error rate. This cannot become a claim that a corresponding percentage of companies was misclassified. Scope, disclosure inequality and intended use remain primary reasons to reconsider public comparison.

Sensitivity also needs source families, case dependence, coding disagreement, missing evidence and operating breadth. Nominal weight can differ from statistical influence when components vary and correlate. Estimating influence in a small selected corpus would not supply correct future weights. [Weights and importance in composite indicators: Closing the gap.](#)

Make institutional trust part of the method

The publication model is deliberate. Inspected BCG Applied AI Index and McKinsey State of AI articles use aggregate respondent comparisons. BCG's Most Innovative Companies 2023 publicly names

and ranks firms using a different reputation/financial-performance composite. These observations establish neither client relationships nor universal consultancy policy. [BCG AI research](#); [McKinsey AI research](#); [BCG innovation methodology](#).

Mayden's accepted direction is public sector research with attributed facts and a future confidential opt-in diagnostic. Previously published data cannot become anonymous. Preserve history and explain the change. Public facts do not imply participation or endorsement.

Appoint a research lead, a consequential-decision reviewer, a correction route and authority to halt unsupported release. The World Bank discontinued Doing Business after data irregularities: institutional standing does not substitute for controls. That case is not causal evidence about Mayden. [World Bank statement](#).

Predeclare intended decisions, units, sampling frame, validation criteria and analysis. Test content, comprehension, coding and role agreement, sector/language comparability and outcomes against simple alternatives. Report attrition and distinguish within-company change from composition or method change. Human appointments, fieldwork and empirical validation remain future gates.

Leadership agenda

Method lead: specify what each item measures and its required evidence.

Release editor: place denominator, period, access limits and assumptions beside each finding.

Programme owner: appoint accountable humans and protect their right to stop publication.

EXHIBIT 02 / CONCEPTUAL TOOL

A reproducible calculation does not validate the construct

Mayden analytical framework / questions to test, not measured maturity stages

01

Computational assurance

Arithmetic replay and deterministic specification tests check calculations. They do not validate the construct.

02

Measurement validation

Human content review, comprehension work, independent coding and fieldwork remain future gates.

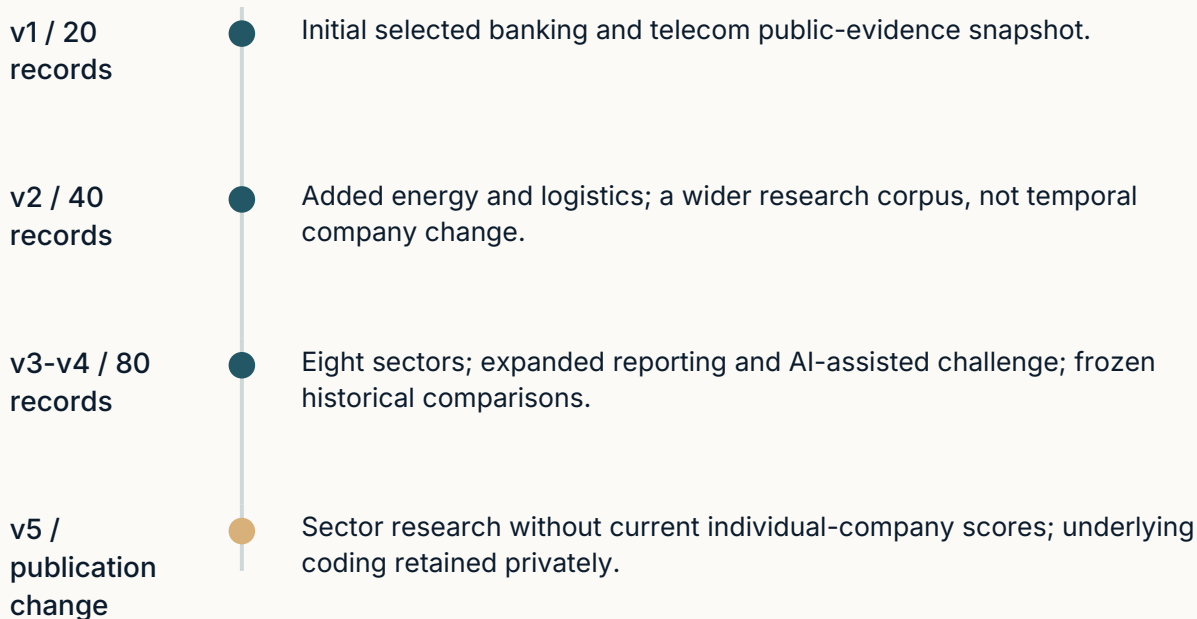
Mayden conceptual framework. These questions specify evidence to request; they do not assert that a named organisation has or lacks the practice. This is not a validated instrument or an observed causal sequence.

Unit: Conceptual decision or evidence request. Research cutoff: 2026-10-10.

EXHIBIT 03 / RESEARCH SYNTHESIS

Historic editions answer historic questions

Mayden analytical framework / questions to test, not measured maturity stages



Source: frozen Mayden edition manifests. Cohort expansion and a publication-policy change do not establish a change inside any company. These are release histories, not a performance time series.

Unit: Frozen publication editions and selected entity records. Research cutoff: 2026-10-10.

EXHIBIT 04 / CONCEPTUAL TOOL

Publish the limits where the result is read

Mayden analytical framework / questions to test, not measured maturity stages

Origin and scope	Who makes the account, and which entity, site or workflow does it concern?
Date and period	When was it published, what period was observed, and is continuing use known?
Unit and denominator	What exactly is counted? Do not mix claims, projects, users, patients or companies.
Boundary and locator	Where is the support, what remains unreported, and which conclusion is not established?

Mayden conceptual framework. These questions specify evidence to request; they do not assert that a named organisation has or lacks the practice. This is not a validated instrument or an observed causal sequence.

Unit: Conceptual decision or evidence request. Research cutoff: 2026-10-10.

05 / THE LEADERSHIP AGENDA

Ask for evidence that changes the decision.

01 / Method lead

Evidence to request: specify what each item measures and its required evidence.

02 / Release editor

Evidence to request: place denominator, period, access limits and assumptions beside each finding.

03 / Programme owner

Evidence to request: appoint accountable humans and protect their right to stop publication.

Method and boundaries

This is an AI-assisted, bounded public-source review of a purposefully selected corpus. It publishes sector research and attributed operating examples, without current individual-company scores, ranks or coded profiles. Source accounts establish what a publisher reports; they do not independently assure operations. The working definition and nine candidate themes require empirical validation.

Recorded source dates run from 2019-05-14 to 2026-10-01. 79 of 224 qualifying claim records have no recorded publication date. This is not a common observation window or confirmation of continuing use.

The confidential opt-in diagnostic remains in development. Independent human review, instrument testing, participant data collection and comparable benchmark norms are future gates. Prior named public-evidence comparisons remain explicitly historical; changes in publication format do not establish changes inside companies.

[Inspect the method](#) / [Explore sector research](#)

06 / SOURCES & PROVENANCE

Follow the argument back to its source.

The following sources support the cited analysis or attributed examples. Inclusion is not independent verification, company endorsement or assurance. Research cutoff: 10 October 2026. Publication revision: 11 October 2026.

S01 / Handbook on Constructing Composite Indicators: Methodology and User Guide

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

Locator: Methodological handbook

[Open source](#)

S02 / Conventional Wisdom on Measurement: A Structural Equation Perspective

Psychological Bulletin / 1991-09 / peer-reviewed-methodological

Locator: Conceptual measurement-model analysis

[Open source](#)

S03 / 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](#)

S04 / PROV-DM: The PROV Data Model

W3C / 2013-04-30 / technical-standard

Locator: Normative data-model specification

[Open source](#)

S05 / Weights and importance in composite indicators: Closing the gap

Ecological Indicators / European Commission JRC / 2017-09 / peer-reviewed-methodological

Locator: Methodological analysis with three case studies

[Open source](#)

S06 / The Formula for Agentic AI Value: Applied AI Index 2026

BCG / 2026-09-30 / Primary institutional publication; provider account, not independent outcome assurance

Locator: Article, Key Takeaways; survey introduction; Sector comparison

[Open source](#)

S07 / The State of AI: Global Survey 2025

McKinsey / 2025 / Primary institutional publication; provider account, not independent outcome assurance

Locator: About the research; exhibits

[Open source](#)

S08 / Most Innovative Companies 2023

BCG / 2023-05 / Primary institutional publication; provider account, not independent outcome assurance

Locator: PDF physical page 7 / printed page 5, Exhibit 3; physical page 27, Methodology

[Open source](#)

S09 / World Bank Group to Discontinue Doing Business Report

World Bank / 2021-09-16 / Primary institutional publication; provider account, not independent outcome assurance

Locator: Official discontinuation statement

[Open source](#)

About Mayden

Mayden connects senior judgement, research and engineering to help organisations turn strategic choices into working capabilities. The AI Native Index is a research programme on recurring work, decisions and accountability.

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