



ASSESSMENT FRAMEWORK

Is your data ready for AI? A scored self-assessment in seven dimensions.

A vendor-neutral way to test whether your data can support a specific AI use case - and to find the one gap that will stall it before you spend on models.

Published

October 2026

Dataddo



AT A GLANCE

One use case, seven dimensions, and a score you can act on.

WHAT YOU DO - 45 TO 90 MINUTES

Score

Pick one concrete AI use case and score 36 statements from 0 to 4 with a data owner and a business owner.

- Seven dimensions, from data access to AI serving
- A five-level maturity scale for every statement
- Evidence required: no proof, no score above 2
- Built on DAMA, NIST AI RMF and the EU AI Act

WHAT YOU GET

Decide

A readiness tier, the gate that caps it, and a 90-day plan matched to where you stand.

- Four tiers: Foundation, Pilot, Production, Scale
- Gating rules so an average cannot hide a missing foundation
- 30, 60 and 90-day actions for each tier
- A vendor checklist if you buy instead of build

Most AI projects stall on data, not on models. **Score the data first, then choose the model.**



Executive summary

Most stalled AI initiatives stall on data, not on models. Gartner predicts that through 2026, organizations will abandon 60% of AI projects that are not supported by AI-ready data, and 63% of organizations say they lack, or are unsure they have, the right data management practices for AI.

This framework gives you a scored, evidence-based view of whether your data can support a **specific** AI use case, and what to fix first. It is built on public, vendor-neutral references: the DAMA data quality dimensions, the NIST AI Risk Management Framework and Article 10 of the EU AI Act (see Sources).

Who it is for: data and analytics leads, CDOs, heads of RevOps or BI, and the business owner of an AI use case.

If your result is...	It means
Foundation	Core data is missing, manual or untrusted. Fix pipelines and ownership before any model work.
Pilot	Data mostly flows, but quality, meaning or controls have gaps. Run time-boxed pilots with a human checking outputs.
Production	Data is reliable, governed and documented for this use case. Go live, with monitoring.
Scale	Data is served to AI through governed, reusable interfaces. Add use cases and agents on the same layer.

Table 1. The four readiness tiers.

How to use the framework

1. **Pick ONE concrete use case**, for example "an AI agent that answers sales pipeline questions". Readiness is use-case specific; a company-wide score hides the gaps that matter.
2. **Score each statement from 0 to 4**, together with at least one data owner and one business owner.
3. **Write down the evidence** behind each score. A score of 3 or 4 without evidence counts as 2.
4. **Compute the scores, apply the gating rules**, and pick actions from the matching tier.

1 The framework at a glance

Readiness is assessed across seven dimensions, each scored on the same five-level scale. The dimensions build on each other: an AI system can only be as good as the least ready layer beneath it. Governance is not a layer of its own; it runs across all of them.

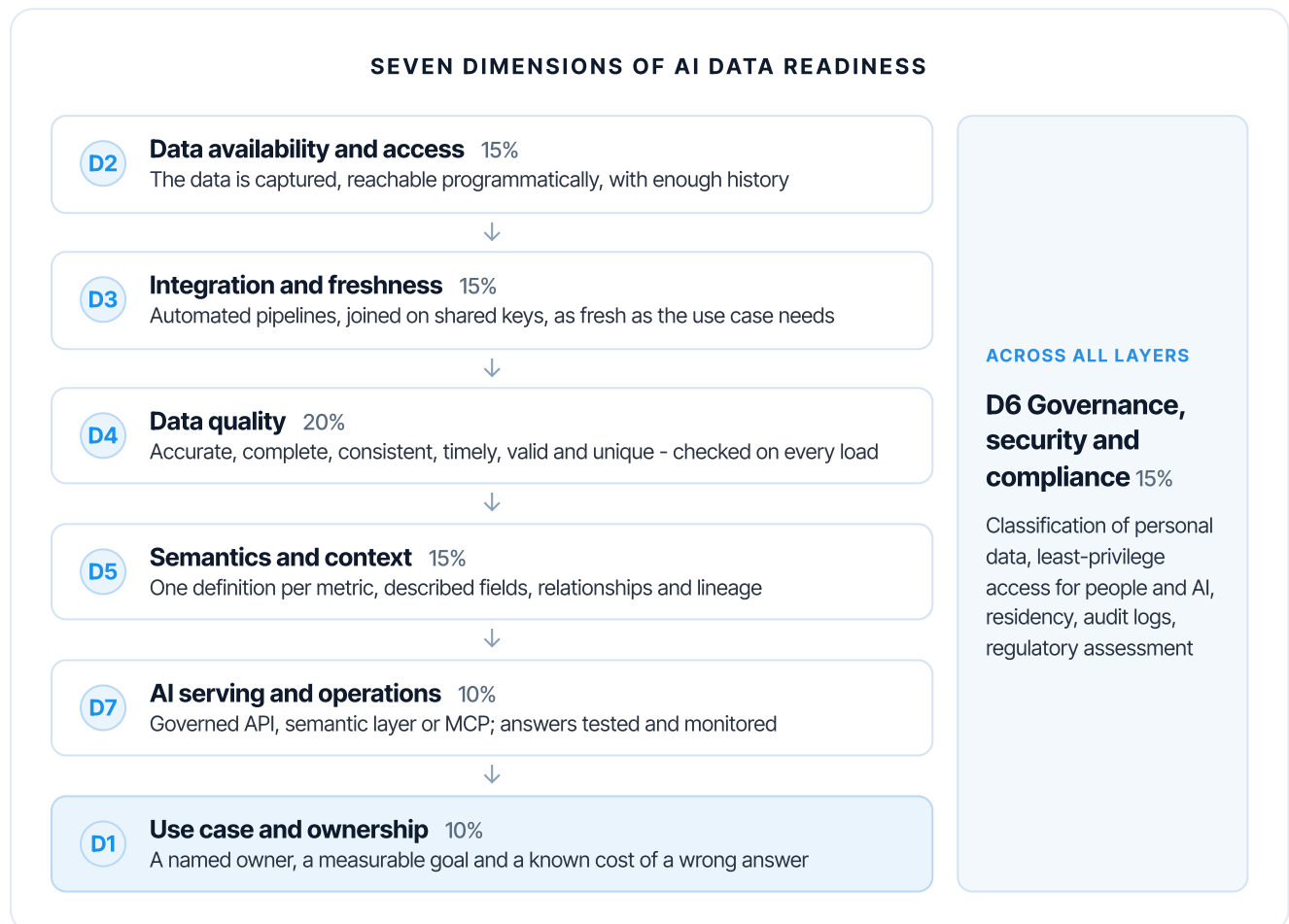


Figure 1. The seven dimensions, read top down: each layer depends on the ones above it. Percentages are the scoring weights.

#	Dimension	The question it answers	Weight
D1	Use case and ownership	Do we know what the AI must do, for whom, and who owns the data behind it?	10%
D2	Data availability and access	Is the data the use case needs captured, and can we reach it?	15%
D3	Integration and freshness	Does the data arrive reliably, combined, and as fresh as the use case needs?	15%
D4	Data quality	Is the data accurate, complete, consistent, timely, valid and unique?	20%
D5	Semantics and context	Would a model (or a new colleague) understand what the data means?	15%
D6	Governance, security and compliance	Do we control who and what can see the data, and can we prove it?	15%
D7	AI serving and operations	Can models and agents consume the data safely, and do we notice when it breaks?	10%

Table 2. Dimensions and weights. Weights reflect the authors' judgment of typical impact on AI outcomes; adjust them for your use case, but agree on them before scoring.

1.1 The maturity scale

Every statement is scored on the same scale.

Score	Level	What it looks like
0	Absent	Not in place, or nobody knows.
1	Ad hoc	Exists in pockets, depends on individuals, manual.
2	Defined	Documented and repeatable, but not measured or enforced.
3	Managed	Owned, measured, monitored, with alerts and SLAs.
4	Optimized	Automated, continuously improved, proven on AI workloads in production.

Table 3. Maturity scale.

2 Assessment questionnaire

Score each statement 0 to 4 for your chosen use case and note the evidence. Statements marked **GATE** are gating: they cap your overall tier (see Section 3).

D1 Use case and ownership 10%

#	Statement	Score (0-4)	Evidence
1.1	The use case has a written goal, a named business owner and a measurable success metric. GATE		
1.2	We have listed the specific data entities and sources the use case needs.		
1.3	Every required source has a named data owner who can approve access and fix issues.		
1.4	We know the cost of a wrong AI answer for this use case (low, medium, high) and have set tolerances accordingly.		

Table 4. D1 Use case and ownership, weight 10%.

D2 Data availability and access 15%

#	Statement	Score (0-4)	Evidence
2.1	All data the use case needs is already captured in a system (not in people's heads, inboxes or slide decks). GATE		
2.2	Every required source, including SaaS apps, ad platforms, databases and files, can be accessed programmatically (API, database, export).		
2.3	Enough history exists for the use case (for example, 12 to 24 months for forecasting).		
2.4	Unstructured sources the use case needs (documents, tickets, call notes) are identified and reachable.		
2.5	Getting access to a new source takes days, not months.		

Table 5. D2 Data availability and access, weight 15%.

D3 Integration and freshness 15%

#	Statement	Score (0-4)	Evidence
3.1	Data from different sources is joined on shared keys (customer, product, account), with a defined way to resolve duplicates.		
3.2	Pipelines run automatically; no step depends on someone downloading or uploading a file. GATE		
3.3	Data freshness matches the use case (for example, hourly for an operational agent, daily for reporting) and is measured.		
3.4	Source API and schema changes are detected and handled without silent breaks.		
3.5	Pipeline failures alert an owner, and failed loads can be re-run or backfilled.		
3.6	The cost of maintaining integrations (people time plus tooling) is known per source.		

Table 6. D3 Integration and freshness, weight 15%.

D4 Data quality 20%

#	Statement	Score (0-4)	Evidence
4.1	Quality rules exist for the critical fields, covering accuracy, completeness, consistency, timeliness, validity and uniqueness.		
4.2	Quality is checked automatically on every load, and results are visible to data and business owners. GATE		
4.3	Known quality issues are logged with an owner and a fix date.		
4.4	We have checked the data for gaps and biases relevant to the use case (for example, missing regions, segments or periods).		
4.5	Business users trust the numbers: the same metric gives the same answer across tools.		

Table 7. D4 Data quality, weight 20%. Source: the six DAMA data quality dimensions, as summarized by IBM and the UK Government Data Quality Hub.

D5 Semantics and context 15%

#	Statement	Score (0-4)	Evidence
5.1	Key business metrics (revenue, active customer, churn, pipeline) have one documented definition. GATE		
5.2	Tables and fields have human-readable names and descriptions, stored in a catalog or semantic layer, not only in people's heads.		
5.3	Relationships between entities are documented, so a model can tell how orders, customers and campaigns connect.		
5.4	We know which datasets are the trusted (certified) version and which are raw or experimental.		
5.5	Lineage shows where each AI-facing field comes from and how it was transformed.		

Table 8. D5 Semantics and context, weight 15%.

D6 Governance, security and compliance 15%

#	Statement	Score (0-4)	Evidence
6.1	Personal and sensitive data in scope is identified and classified. GATE		
6.2	Access is role-based and least-privilege, and the same rules apply when an AI model or agent queries the data. GATE		
6.3	We know where data is stored and processed (region, vendor) and it meets our residency and contractual obligations.		
6.4	Every AI access to data is logged and auditable.		
6.5	We have assessed whether the use case falls under regulation such as the EU AI Act, and documented data origin, preparation and assumptions.		
6.6	Data vendors and processors in the chain hold relevant certifications (for example, SOC 2 Type II, ISO 27001).		

Table 9. D6 Governance, security and compliance, weight 15%. Sources: EU AI Act, Article 10 for statement 6.5; NIST AI RMF for governance practices.

D7 AI serving and operations 10%

#	Statement	Score (0-4)	Evidence
7.1	Models and agents get data through a governed interface (API, semantic layer, MCP server), not through copied extracts.		
7.2	AI tools can access only the data needed for the use case, scoped per user or role, read-only where possible.		
7.3	AI answers are evaluated against a test set of known-correct questions before and after changes.		
7.4	Data issues found through AI usage (wrong answers, missing fields) flow back to data owners.		
7.5	We can switch AI models or vendors without rebuilding the data layer.		

Table 10. D7 AI serving and operations, weight 10%.

3 Scoring and interpretation

Your readiness tier is the lower of two results: the weighted score, and the cap set by the gating statements. A high average cannot compensate for a missing foundation.

3.1 Compute the scores

STEP 1 - DIMENSION SCORE (0 TO 100)

$$D_i = (\text{average statement score in the dimension} \div 4) \times 100$$

STEP 2 - OVERALL SCORE (0 TO 100)

$$\text{Overall} = 0.10 D_1 + 0.15 D_2 + 0.15 D_3 + 0.20 D_4 + 0.15 D_5 + 0.15 D_6 + 0.10 D_7$$

3.2 Apply the gates

Two checks run before the score is read. An overall score of 80 with a manual pipeline (statement 3.2 scored 1) is still Foundation.

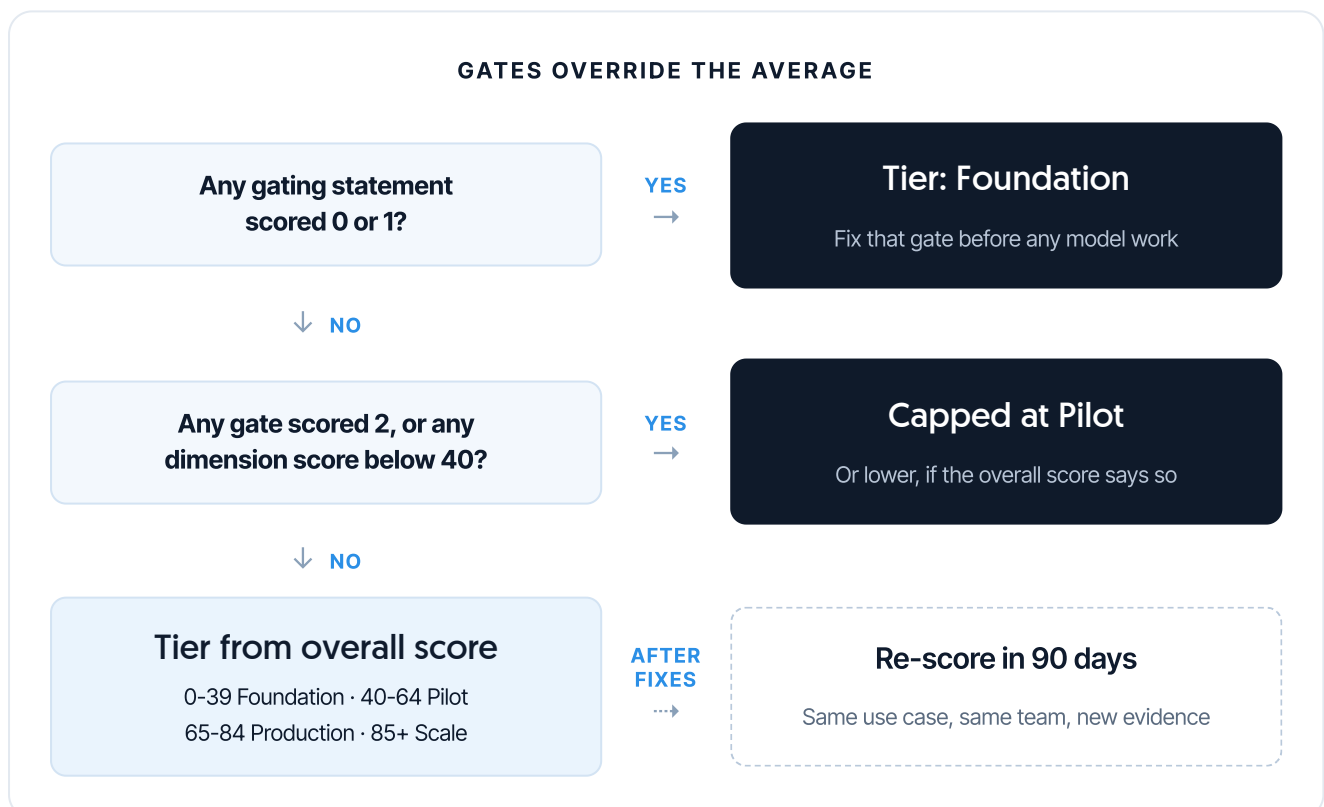


Figure 2. Gating rules: two checks, three outcomes, one re-assessment loop.

3.3 Read your tier

Tier	Overall score	What it means	What AI you can run safely
Foundation	0 to 39	Core data is missing, manual or untrusted.	Internal experiments on sample data only.
Pilot	40 to 64	Data exists and mostly flows, but quality, meaning or controls have gaps.	Time-boxed pilots with a human checking every output.
Production	65 to 84	Data is reliable, governed and documented for this use case.	Production use for this use case, with monitoring.
Scale	85 to 100	Data is served to AI through governed, reusable interfaces.	Several use cases and autonomous agents on the same data layer.

Table 11. Readiness tiers.

Worked example. A marketing team scores D1 = 75, D2 = 70, D3 = 45, D4 = 60, D5 = 40, D6 = 70, D7 = 50. The overall score is 58.25, which is Pilot. Statement 3.2 (manual CSV uploads from two ad platforms) scored 1, so the gate caps them at **Foundation**: automating those pipelines is the first fix, before any model work.

4 Action plan by tier

Fix the gate that caps your tier first, then the dimension with the lowest score. Each tier below lists what to do in the next 90 days.

Tier	First 30 days	Days 31 to 60	Days 61 to 90
Foundation	Narrow to one use case with an owner and a metric. Inventory required sources and owners.	Replace manual exports with automated pipelines for the required sources. Centralize data in one store (warehouse, lakehouse or database).	Add basic quality checks and failure alerts on those pipelines. Re-run this assessment.
Pilot	Write one definition for the 5 to 10 metrics the use case depends on. Classify sensitive fields.	Add descriptions to AI-facing tables and fields. Set freshness targets and measure them.	Build a test set of 30 to 50 known-correct questions. Run a supervised pilot.
Production	Put AI access behind a governed, read-only interface with role-based scopes. Turn on audit logging.	Monitor quality, freshness and answer accuracy together, with owners on alert.	Document data origin, preparation and assumptions for compliance. Plan the second use case on the same layer.
Scale	Certify reusable datasets and publish them in a catalog.	Make the data layer model-agnostic, so models and agents can be swapped.	Track cost per use case (pipelines, storage, tokens) and retire unused data.

Table 12. 90-day actions by readiness tier.

4.1 Build or buy the data plumbing?

Teams in the Foundation and Pilot tiers usually spend most of their effort on D2 and D3: connecting sources and keeping pipelines alive. Decide deliberately whether that is where your engineers should spend their time.

Factor	Favors building in-house	Favors a managed integration platform
Sources	A few internal databases	Many SaaS apps, ad and CRM platforms with changing APIs
Team	Dedicated data engineers with spare capacity	Small data team, analysts or RevOps owning data
Change rate	Stable schemas	Frequent API and schema changes
Compliance	Custom controls you already run	Need for certified vendors, residency options and audit logs out of the box
Time to value	Months are acceptable	Weeks or less

Table 13. Build or buy. Whichever route you take, measure total cost: engineering hours to build and maintain each connector, plus tooling and incident time.

5 Evaluating data integration vendors for AI

If you buy rather than build, ask every shortlisted vendor the same questions and ask for proof, not slides. The questions map to the dimensions above.

Area	Question to ask	Proof to request	Maps to
Coverage	Are all our required sources and destinations supported by maintained connectors? What happens for one that is not?	Connector list; SLA for new connector builds	D2
Maintenance	Who fixes a connector when a source API changes, and how fast?	Change-handling process; incident history	D3
Reliability	How are failures, retries and backfills handled and reported?	Monitoring and alerting demo	D3, D7
Quality	Can we validate and monitor data on load, before it reaches models?	Quality rules demo	D4
Semantics	Can we expose business definitions and descriptions alongside the data?	Semantic or metadata layer demo	D5
Security	Which certifications do you hold? How are credentials stored?	SOC 2 Type II report, ISO 27001 certificate	D6
Residency	Where is data processed and stored? Can we run it in our own cloud or on-premise?	Region list; deployment options	D6
AI access	How do AI models and agents access the data: API, MCP, semantic layer? Is it read-only and scoped per user?	Live demo with an AI assistant	D7
Portability	Can we switch destinations, warehouses or AI models without rebuilding pipelines?	Reference customers who have done it	D7
Cost	How does pricing scale with sources, rows and users? What would our cost be in 2 years at 3x volume?	Written quote at current and 3x volume	All

Table 14. Vendor evaluation questions.

- ☐ Shortlist no more than three vendors.
- ☐ Run a proof of concept on your hardest source, not your easiest.
- ☐ Score the proof of concept with this framework's D2, D3 and D7 statements.

6 About Dataddo and sources

This framework is published by Dataddo, a data integration platform with 400+ managed connectors that moves data between sources and destinations and serves governed business data to LLMs and agents through UI, API and MCP. Dataddo runs in cloud, on-premise and hybrid deployments and holds SOC 2 Type II and ISO 27001 certifications.

If your assessment shows gaps in D2 (access), D3 (integration and freshness) or D7 (AI serving), those are the dimensions Dataddo is built to close.

This framework was produced by Dataddo. The dimensions, weights and tier thresholds are the authors' own and are not taken from any standard. Product capabilities and certifications are as published on dataddo.com in October 2026 and may change.

Close your data gaps before your AI goes live

Connect the sources your use case needs and serve governed data to Claude, ChatGPT and your own agents - or plug Dataddo into the warehouse, lake or lakehouse you already run.

Try Data to AI™

Explore Data Anywhere™

Sources

All pages accessed October 24, 2026.

- [Gartner, Lack of AI-Ready Data Puts AI Projects at Risk, February 2025](#)
- [IBM, Data quality dimensions \(DAMA\)](#)
- [UK Government Data Quality Hub, Meet the data quality dimensions](#)
- [NIST, AI Risk Management Framework](#)
- [EU AI Act, Article 10: Data and data governance](#)
- [Dataddo](#)