

Arx Certa

Expert UK Cloud, Infrastructure & AI

AI READINESS · DATA

AI Data Readiness Checklist

12 questions to ask before any AI project touches your data.

Free download · May 2026 · v2026-05

Scorecard companion: arxcerta.com/ai-readiness-scorecard

Why data readiness is the silent killer

Most AI projects don't fail because the model was wrong. They fail because the data was wrong — incomplete, inconsistent, mis-classified, or out of reach behind systems nobody had time to integrate. The pilot ships, the outputs disappoint, the vendor gets blamed, and the real culprit — the data layer — never makes it into the post-mortem.

"*The data will be fine*" is the most expensive sentence in AI procurement. These twelve questions are the conversation that should happen **before** a vendor walks through the door.

How to use this. Score each question 1–3. 1 = no, 2 = mostly, 3 = yes with evidence. Aggregate matters less than the shape: any 1 in questions 1–4 means data work has to lead the AI work.

The 12 questions

1. Where does the data we want our AI to use actually live? (System, owner, format.)
2. Can we connect to that source via an API, today? (Or do we need a custom integration first?)
3. Is the data structured or unstructured? (Most useful AI use cases need both, blended.)
4. How clean is it? (Duplicate records, missing fields, inconsistent formats.)
5. How current is it? (Real-time, daily, weekly, or quarterly?)
6. How is it secured today? (Who has access, how is access logged, who can revoke it?)
7. Do we have lineage? (Where each field came from and when it last changed.)
8. Have we classified it by sensitivity? (Public / business / confidential / restricted.)
9. Do we have a retention policy that AI use will need to respect?
10. Have we ever had a data quality incident? What was the root cause and what changed afterwards?
11. Do we have a single named owner for this data?
12. If we built an AI tool on this data today, what's the worst thing it could output — and would we catch it before it left the building?

How to use this

Run this once with the data owners in the room, not from a desk. Score each question 1–3 with at least one piece of evidence behind every score above 1. The exercise itself surfaces the work that needs to happen before any AI vendor conversation.

Scoring matrix

Score	What it means	What to do
1 — No	We can't answer the question, or the answer is no.	Block this AI use case until resolved.
2 — Mostly	We can answer partially or for some data, not all.	Acceptable for a contained pilot; resolve before scale.
3 — Yes, with evidence	We can demonstrate the answer with documentation, dashboards, or audit logs.	Green for AI work.

Threshold

Any score of 1 on questions 1–4 (location, integration, structure, cleanliness) means data foundations work needs to happen **before** AI implementation work. Any score of 1 on questions 6–8 (security, lineage, classification) means governance work needs to run in parallel with the AI work, not after it.

Want a full picture, not just the data dimension?

The AI Readiness Scorecard scores you across five dimensions — Governance, Data, Infrastructure, Security, and Use case — in 4 minutes. Free, with a personalised 30-day action plan.

Take the scorecard → arxcerta.com/ai-readiness-scorecard