

BEYOND THE ALGORITHM

Data Readiness as AI Prerequisite — The Boring Work Nobody Wants to Fund

A Practitioner's Perspective on the Gap Between AI Ambition and Data Reality

Giovanni Leonardi

July 2022



Data Readiness as AI Prerequisite — The Boring Work Nobody Wants to Fund

Transformation · No. 281

Originally written July 2022 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“The organisations that will succeed with AI are not the ones with the most sophisticated models — they are the ones that did the unglamorous work of making their data findable, usable, and trustworthy before the pressure to deploy became irresistible.”

1. The Investment Nobody Wants to Make

There is a pattern I have observed with striking consistency across sectors and organisation sizes over the past several years. A board or executive committee becomes convinced — rightly — that artificial intelligence and machine learning represent a strategic imperative. A budget is allocated. A team is assembled or a consultancy engaged. Ambitions are declared. And then, within months, the initiative stalls — not because the algorithms are wrong, or the use cases poorly chosen, but because the data is not ready.

The reasons are always similar. Data sits in silos that were never designed to interoperate. Definitions vary between departments — what finance calls a customer is not what marketing calls a customer. Quality is inconsistent, with gaps, duplicates, and stale records accumulated over years of under-investment in stewardship. Metadata is sparse or absent. Lineage is untraceable. And governance, where it exists at all, is a compliance exercise rather than an operational discipline.

This is not a new problem. But what makes it newly urgent is the gap between the sophistication of the AI ambition and the immaturity of the data foundation it depends on. Organisations are trying to build predictive models on data estates that cannot reliably produce an accurate management report.

2. Why Data Readiness Is Systematically Underfunded

The root cause is not ignorance. Most senior leaders, when pressed, will acknowledge that data quality matters. The problem is structural: data readiness work is invisible, unglamorous, and difficult to attach to a compelling business case in isolation.

Consider what data readiness actually involves. It means cataloguing what data exists and where it lives. It means reconciling conflicting definitions across business units that have operated independently for years. It means establishing ownership — persuading people

to accept accountability for data assets they have historically treated as someone else's problem. It means building pipelines that clean, validate, and transform data reliably, not as a one-off project but as an ongoing operational capability. And it means creating governance structures that are lightweight enough to be followed and rigorous enough to be meaningful.

None of this produces a demonstration that excites a steering committee. There is no model to show, no prediction to marvel at, no proof of concept to circulate. The output is plumbing — essential, but invisible when it works and noticed only when it fails.

The result is a predictable funding pattern. The AI initiative receives a substantial budget, often anchored to a high-profile use case. The data work is either assumed to be a minor prerequisite — a few sprints of preparation before the real work begins — or it is scoped separately as a "data quality programme" that must compete for funding on its own merits, without the strategic urgency that AI carries.

In my experience, the data work typically needs sixty to seventy per cent of the total effort in the first year of any serious AI programme. It routinely receives less than twenty.

3. The Consequences of Getting This Wrong

The consequences are not abstract. They manifest in specific, recurring failure modes.

The first is the proof-of-concept trap. A team builds a working model on a curated sample dataset — clean, well-structured, carefully prepared. The demonstration succeeds. Approval is given to scale. And then the model meets production data: messy, incomplete, inconsistent, arriving at unpredictable intervals through brittle integrations. Performance degrades. The team spends months debugging data issues rather than refining the model. The timeline slips. Confidence erodes.

The second is governance debt. Without clear ownership and lineage, it becomes impossible to answer basic questions about the data feeding a model. Where did this data come from? When was it last updated? What transformations were applied? Who approved its use for this purpose? In regulated industries — financial services, healthcare, energy — these are not optional questions. They are the questions an auditor will ask, and the inability to answer them can halt a deployment entirely.

The third is silent bias and error propagation. When data quality is poor and governance is weak, errors and biases in source data flow unchecked into model training. The model learns the biases, amplifies them, and presents the results with the false confidence of algorithmic authority. The organisation then makes decisions on the basis of outputs that are systematically skewed, without the visibility to recognise the distortion.

4. What Ready Actually Looks Like

Data readiness is not a binary state. It is a spectrum, and the level required depends on the ambition. But there are baseline capabilities without which any AI initiative is building on sand.

- A data catalogue that reflects reality. Not a one-off inventory that was accurate eighteen months ago, but a living, maintained view of what data exists, where it lives, who owns it, and what it means.
- Agreed definitions. When the model says "active customer" or "monthly revenue" or "product category," every stakeholder understands the same thing. This sounds trivial. In practice, it is one of the hardest problems in enterprise data management.
- Measurable quality. Not a vague aspiration to "improve data quality," but defined metrics — completeness, accuracy, timeliness, consistency — measured continuously and reported transparently.
- Operational governance. Roles and responsibilities that are clear, accepted, and enforced. Data owners who actually own decisions about their data. Stewards who have the mandate and the capacity to maintain standards.
- Repeatable pipelines. The ability to extract, transform, and load data reliably, with logging, error handling, and monitoring — not as a project deliverable but as an ongoing operational service.

The organisations that get this right share a common characteristic: they treat data readiness not as a precursor to the AI programme but as part of it. The data work is on the same roadmap, funded from the same budget, governed by the same steering group, and held to the same delivery standards. It is not a separate stream that must justify itself independently.

5. The Leadership Challenge

Ultimately, this is a leadership problem more than a technical one. The technology to catalogue, clean, govern, and pipeline data is mature and widely available. What is missing is the organisational will to prioritise work that is necessary but unexciting.

This requires leaders who are willing to tell the truth about where their organisation actually stands, rather than where it would like to be. It requires honest assessment — the kind that acknowledges a data estate is not AI-ready, even when the board has already announced an AI strategy. And it requires the patience to invest in foundations before pursuing the ambitions that those foundations make possible.

The organisations that will succeed with AI are not the ones with the most sophisticated models — they are the ones that did the unglamorous work of making their data findable, usable, and trustworthy before the pressure to deploy became irresistible.

In the current climate, where every sector is racing to demonstrate AI capability, this is an unpopular message. But it is, in my observation, the single most important determinant of whether an AI investment delivers lasting value or joins the growing list of initiatives that promised transformation and delivered disappointment.

The boring work is the real work. The question is whether enough leaders have the courage to fund it.