TRANSFORMATION BOOK SERIES · No. 349

White Paper

BEYOND MODEL AMBITION

AI Strategy Fails at the Data Layer — and That Is the Work Nobody Wants to Fund

Evidence from early enterprise AI programmes points to one recommendation: treat data readiness as a governed product, not a pre-project clean-up

Giovanni Leonardi

August 2022

AI Strategy Fails at the Data Layer — and That Is the Work Nobody Wants to Fund

Transformation · No. 349
Originally written August 2022 | Re-edited and re-rendered for the WhereBeyond Edition,

"AI does not remove the cost of poor data; it converts that cost into hidden model risk."

1. AI Strategy Fails at the Data Layer — and That Is the Work Nobody Wants to Fund

Evidence from early enterprise AI programmes points to one recommendation: treat data readiness as a governed product, not a pre-project clean-up

An executive team approves an artificial-intelligence pilot to improve the triage of service cases. The model is expected to reduce manual review, identify urgent cases earlier and demonstrate visible progress within twelve weeks. The technology is available. A specialist team is appointed. The initial business case is persuasive.

In week three, the team discovers that "urgent" has four operational definitions. In week five, it finds that 14 per cent of historical cases lack a reliable closure code. In week seven, two source systems produce different customer identifiers. By week ten, most of the pilot budget has paid skilled data scientists to reconcile records and negotiate definitions.

The model has not failed. The organisation has asked it to learn from ambiguity that the organisation itself has never resolved.

This pattern is recurring across enterprise AI programmes. Investment gravitates towards models, platforms and visible demonstrations because they look like innovation. Data readiness is treated as preparation: important, certainly, but assumed to be a technical task that can be completed cheaply before the real programme begins.

That assumption is wrong. Data readiness is not the cleaning of a dataset. It is the establishment of durable organisational conditions under which data can be understood, accessed, trusted and changed for a defined purpose. It requires ownership, operational decisions, controls and sustained funding. Without those conditions, AI pilots may still produce promising demonstrations, but they do not produce dependable operating capability.

AI does not remove the cost of poor data; it converts that cost into hidden model risk.

2. The evidence sits inside failed pilots

The most useful evidence is often found not in model accuracy scores but in how project time is actually spent.

Across recognisable enterprise programmes, the same sequence appears:

- a use case is chosen for apparent value and data availability
- historical data is extracted and found to contain undocumented operational variation
- specialists repair the sample sufficiently to train a model
- the pilot performs acceptably under controlled conditions
- deployment exposes new data, process and ownership problems
- confidence falls, exceptions rise and the model remains in supervised use
- the pilot is described as technically successful but difficult to scale

This is not a minor implementation problem. It reveals that the demonstration and the operating service are drawing on different data conditions.

A pilot can rely on a static extract, hand-labelled records and direct access to knowledgeable staff. An operating model must cope with changing source systems, new categories, inconsistent entry, access restrictions, delayed corrections and shifts in customer or employee behaviour. If the programme has funded only the model, the conditions that made the pilot credible disappear at deployment.

The evidence is also visible in the composition of effort. A notional sixteen-week pilot may allocate six weeks to model development but consume ten weeks in locating fields, explaining codes, resolving access and rebuilding extracts. This work is reported as delay because the plan regarded data as an input. In reality, it is the discovery of the organisation's unpriced data estate.

The correct conclusion is not that AI is immature. It is that the business case was incomplete.

3. What "ready" actually means

Data readiness is frequently reduced to quality: completeness, accuracy, timeliness and consistency. Those measures matter, but they are not sufficient. A dataset can be technically clean and still be unfit for an AI-enabled decision.

Readiness has at least five dimensions.

Dimension	Evidence of readiness	Typical failure
Meaning	Critical fields and outcomes have agreed operational definitions	Different functions use the same label for different events
Provenance	The source, transformation and limitations of data can be traced	A model team cannot explain how a field was derived
Fitness	Data represents the population and decision the use case concerns	Historical records reflect a process that is now changing
Stewardship	Named owners can authorise access and resolve defects	Every issue is acknowledged but no one can decide
Continuity	Quality and drift can be monitored after deployment	A one-off cleaned extract becomes the unofficial production design

These dimensions expose why data work is organisational rather than merely technical.

Meaning requires operational leaders to agree what the business is measuring. Provenance requires architecture and engineering discipline. Fitness requires judgement about bias, coverage and changing behaviour. Stewardship requires authority. Continuity requires funding beyond the pilot.

A programme that asks a data team to solve all five is delegating business decisions to people who do not own their consequences.

4. A composite case: the £1.2 million demonstration

Consider a composite organisation seeking to predict which customer applications will require additional review. It has 1.8 million historical records across three systems and expects the model to reduce handling effort by 20 per cent.

The first sample appears encouraging. After excluding incomplete records, the model identifies a large proportion of complex cases. The programme funds a £1.2 million pilot, including integration, specialist support and operational testing.

The apparent evidence weakens under examination.

The excluded records are not random. They contain a disproportionate number of cases from one channel and from customers whose applications required manual intervention. The outcome field records whether a case was referred, but referral practice changed eighteen months earlier. One operational team used referral to manage workload; another used it only when additional evidence was needed. The model is therefore learning both customer complexity and historical staffing practice.

The programme has three choices.

1. Proceed quickly. Deploy the model as decision support, retain human review and accept lower benefits while performance is observed.
2. Repair the dataset. Reconstruct definitions, relabel a representative sample and delay deployment.
3. Redesign the use case. Narrow the model to a decision supported by more reliable data, sacrificing some headline value for greater control.

The first option protects momentum but transfers cost into supervision. The second can improve the model but may create a bespoke historical dataset that is expensive to maintain. The third reduces ambition but aligns the use case with evidence the organisation can sustain.

The programme chooses the second option. A cross-functional team reviews 40,000 records, defines six referral categories and establishes a common identifier. Model performance improves. Yet six months later, a source-system change alters two key fields and the relabelling process has no funded owner. The data was made ready for the project, not kept ready for the service.

The lesson is not that the repair was wasted. It is that readiness was treated as a finite deliverable when it was an operating capability.

5. The strongest opposing view

A serious objection is that this argument risks making AI programmes impossibly cautious. Enterprise data has never been perfect. If every definition, lineage issue and ownership gap must be resolved before experimentation, organisations will spend years on data programmes and learn nothing about where AI creates value.

There is additional force to the objection in 2022. Pre-trained models, cloud platforms and packaged services are reducing some of the need to build models from the ground up. Transfer learning can lower the volume of organisation-specific training data

required. Synthetic or externally sourced data may help in certain settings. Rapid experiments can reveal defects that a broad data-governance initiative would never prioritise.

This is correct. “Fix the data first” is poor advice if it means cleanse the entire estate before selecting a use case. Large, untargeted data programmes often create catalogues, standards and repositories without changing a single operational decision. They can consume substantial investment while remaining detached from value.

But it does not follow that data readiness can be deferred. It means readiness must be use-case-led and proportionate.

The requirement is not perfect data. It is explicit knowledge of the data’s limitations, accountable acceptance of the resulting risk and a mechanism for maintaining fitness as the service changes. Experimentation is valuable when it discovers those conditions early. It becomes wasteful when the same discovery is repeated in every pilot because nobody funds the shared capability.

The choice is therefore not between years of data preparation and reckless model deployment. The choice is between governed discovery and recurring surprise.

> *Fund the minimum data capability required to operate the use case safely—not the maximum data estate that could someday be useful.*

6. Why the business case keeps excluding the real work

Data readiness struggles for funding because its benefits are indirect and shared.

A model can be attached to a visible outcome: fewer cases reviewed, faster decisions, lower loss or better forecasting. A common definition, lineage record or stewardship process supports several outcomes without owning any one of them. Traditional investment governance then asks each project to justify only its local scope.

The incentives produce three distortions.

- Preparation is hidden inside delivery. Model specialists become expensive data-repair capacity.
- Shared capability is repeatedly rebuilt. Each pilot creates its own extracts, labels and definitions.

- Maintenance is omitted. The business case funds accuracy at launch but not monitoring, correction or change control.

The accounting looks efficient because the enabling work is fragmented. The portfolio is inefficient because it pays for the same uncertainty several times.

There is also a governance problem. Data owners are frequently named without being given time, authority or measures. An owner who cannot settle a definition, require a source-system correction or approve a justified exception is not an owner. It is a contact name.

7. Recommendation: fund readiness as a portfolio of data products

The more effective response is to organise and fund data readiness as a set of governed data products tied to priority AI use cases.

A data product in this context is not simply a technology component. It is a maintained body of data, meaning, controls and service commitments with a named owner and known consumers. Its purpose is to make a defined class of decisions supportable.

The portfolio should proceed in four moves.

1. Select decisions before models. Define the operational decision, its value, the consequence of error and the human accountability that remains. This prevents an attractive technique from searching for a problem.

2. Assess the decision's data contract. Specify required fields, definitions, provenance, access, acceptable gaps, refresh frequency and monitoring. Record what is known, not what the business case wishes were true.

3. Fund the shared readiness gap. Separate reusable work—identifiers, definitions, lineage, access patterns and stewardship—from model-specific preparation. Allocate shared work at portfolio level so it is not repeatedly cut by individual projects.

4. Gate deployment on operating evidence. Require evidence that the live data flow, exception process, ownership and monitoring work under realistic conditions. A high offline accuracy score is not a deployment decision.

Each data product needs a compact set of roles:

- an operational owner accountable for meaning and use
- a data steward responsible for issue resolution and quality rules
- an engineering owner accountable for pipelines, lineage and service performance
- a model owner accountable for performance, limitations and monitoring

- a risk owner able to accept or reject the residual consequence

These roles can be combined in small organisations, but the accountabilities cannot be omitted.

8. The investment test should change

Approval forums should stop asking only whether sufficient data exists to begin. That question encourages optimistic answers and one-off extracts.

They should ask:

- What decision will this data support?
- Which limitations could change or invalidate that decision?
- Who can resolve a definition when functions disagree?
- What happens when the source process changes?
- Which readiness work is reusable across the portfolio?
- What continuing cost keeps the data fit after deployment?
- What evidence would cause the organisation to suspend or narrow the model?

These questions make the hidden work visible before it becomes sunk cost.

They also improve prioritisation. A high-value use case with severe data uncertainty may still proceed, but first as a governed discovery. A modest use case with strong data foundations may reach operation quickly and establish reusable capability. A glamorous use case without an accountable decision owner should not receive funding merely because the model can be demonstrated.

9. The unglamorous foundation is the strategic choice

AI investment is likely to continue accelerating. More capable models and more accessible platforms will make experimentation easier. That will increase, not reduce, the importance of data readiness. When model construction becomes less scarce, the organisation's distinctive advantage shifts towards the quality of its operational context: what its data means, how quickly it can be trusted and whether it can be maintained as conditions change.

The boring work is boring because it forces decisions that technology cannot make on the organisation's behalf. What is a customer? Which outcome counts as success? Who owns the correction? How much error is acceptable? Which process will change when the

evidence changes?

Those questions do not produce impressive demonstrations. They produce dependable capability.

The recommendation is therefore direct: do not fund data readiness as preliminary cleaning, and do not launch a universal data programme detached from use cases. Fund a governed portfolio of data products, each anchored to real decisions and maintained as an operating service.

That approach does not promise perfect data. It creates something more valuable: data whose imperfections are understood, owned and controlled well enough for AI to be used with confidence.

The organisations that do this work will appear slower at the beginning because they are pricing the whole system. They will move faster after the pilot because they will not have to rediscover, reconstruct and renegotiate the foundation every time.