

85% of AI Projects Fail to Deliver Value.

47% Fail Because the Data Underneath Is Broken.

Fix your data before you buy the AI. Here's the four-step way companies do it right.

Why Data Comes Before AI

01 Three ways to fix scattered data

Lakehouse, Data Mesh, or Data Fabric — each trades speed against cost.

02 Four steps before AI goes live

Find the silos, set clear rules, prep the text, check quality automatically.

03 Real proof it works

JPMorgan Chase and Netflix both fixed data first — and moved faster, not slower.

47%

of AI project failures trace straight back to fragmented data.

Data scattered across old, disconnected systems — not weak AI models. —
Gartner, 2023

DATA FRAGMENTATION

2.5x

more likely to get real value from AI when data is unified.

Companies with one connected data platform beat companies with scattered data. — McKinsey & Company, 2023 AI Survey

DATA FRAGMENTATION



Buying a powerful AI model before fixing scattered data is like installing a smart thermostat in a house with no wiring.

The technology is good. It has nothing solid to connect to.

DATA FRAGMENTATION

Three Ways to Fix Scattered Data

APPROACH	TIME TO BUILD	BIGGEST RISK
Centralised Lakehouse	12–18 months	Expensive; can become a new bottleneck
Data Mesh	9–15 months	Needs a real culture change
Data Fabric	3–6 months	Slower on live AI search

Four Steps Before Any AI Goes Live

01 Find every data silo

Scan every database, tag private data, trace where it all goes.

02 Set clear ownership rules

One owner per data type, shared naming, role-based access.

03 Prep text for AI search

Clean documents, chunk them, store them in a search-ready database.

04 Check quality automatically

Reject bad records, watch for drift, flag duplicates instantly.

Two Companies That Fixed Data First

JPM JPMorgan Chase — weeks to hours

3,000+ pipelines moved to a Data Mesh; data search time fell from weeks to hours, enabling AI tools for 200,000 employees.

NFX Netflix — 40% faster setup

One shared metadata catalog let teams share petabyte-scale data without copying files.

Three Risks to Manage

01 Data leakage

Weak access controls can expose restricted files to the wrong employees.

02 Compliance fines

Rules like the EU AI Act require clear proof of where training data came from.

03 Storage cost

Moving every old file to the cloud gets expensive fast — virtualise low-priority data instead.

Fix the Foundation. Then Build.

The order of operations decides whether AI actually pays off, or just adds to next year's clean-up bill.



Terence Kok · AI & Innovation Strategist

[SAVE](#) · [COMMENT](#) · [FOLLOW](#)