

Data Quality In The Age of Generative AI

A data enablement business, built to turn your data investment into data capability.

NEPHOS



Agenda

01

The AI Impact

02

AI Raises The Stakes

03

Closing The Gap

04

Some Practical
Steps



Most organisations are building AI on data they don't trust

LIMITING FACTOR

52%

of organisations cite data quality as their
#1 barrier to AI adoption

PEX Network, 2025/26 · 200+ professionals surveyed

COST OF FAILURE

\$12.9M

average annual cost of poor data quality
per organisation

*Gartner · IBM estimates \$3.1 trillion across the US
economy*

FAILURE RISK

60%

of AI projects lacking AI-ready data
will be abandoned by 2026

*Gartner · 50% of GenAI POCs already abandoned post-
pilot*

The data problem isn't theoretical. It's quantified. And it's getting more expensive as AI spend accelerates.

Why GenAI Changes The Data Quality Stakes

01

Garbage In, Hallucination Out

- AI models amplify data quality problems, they don't absorb them
- They don't flag low quality inputs
- **They use poor quality data, at speed and with confidence**

02

RAG Is Only As Good As The Corpus

- RAG is the dominant enterprise AI pattern
- If your retrieval layer pulls stale, duplicate, or miss-labelled data — the answer is wrong and sounds authoritative.

03

Legal Liability

- An AI generated response using poor data isn't a product failure, it's a compliance event
- DORA, EU AI Act all point back to data quality and data lineage

Data Quality We Know And Love



The 6 Classic Dimensions

Traditional World

Completeness

Gaps are visible — nulls appear in dashboards



Accuracy

Errors are contained — wrong number in one report



Consistency

Causes internal friction — teams disagree on numbers



Timeliness

Stale data has visible caveats — reports are dated



Uniqueness

Duplicates cause reporting errors — inflated counts



Validity

Invalid data breaks pipelines — throws errors



AI World

Gaps are hidden — models fill them with confident inference

Errors are multiplied — one bad pattern scales across thousands of outputs

Causes external confusion — same question, different answers

Stale data has false authority — AI answers carry no timestamp

Duplicates cause identity confusion — blended profiles at delivery

Invalid data bends outputs — silently absorbed and pattern-matched

Two New Dimensions For GenAI



7 — Lineage

Traditional: Nice-to-have — useful for audit and debugging

→ AI Era: Non-negotiable — regulators require you to explain what data underpinned an AI-generated decision

"If you can't trace an AI output back to its source data, you can't defend it to a regulator."



8 — Semantic Clarity (requires Discovery & Classification)

Traditional: Metadata and taxonomy were governance nice-to-haves

→ AI Era: Models interpret labels and descriptions. Inconsistent taxonomy produces inconsistent model behaviour — at scale.

The prerequisite: Discovery & Classification

You can't have semantic clarity without first knowing what your data is. Discovery tells you what exists across your estate — structured, unstructured, shadow. Classification assigns meaning: sensitivity, type, owner. Without these, your metadata is incomplete and your AI is operating blind.

The Practical Fix – What Needs To Change

1

Discover, Classify & Label Before You Deploy

You can't trust data you don't understand. Map your estate, classify sensitivity and type, assign ownership. This is the foundation everything else sits on.

Discover First

2

Treat Your Data Catalogue, Like AI Infrastructure

Not a governance nice-to-have. If your data isn't described, classified, and trusted — your AI can't use it safely.

Catalogue

3

Shift Quality Checks Left

Into the pipeline, before data reaches the model layer. Quality at ingestion, not remediation after the fact.

Prevention

4

Instrument For Explainability

Know where every AI output came from. Data lineage isn't just audit trail — it's how you defend an AI-generated decision to a regulator.

Lineage

There Is An Execution Gap In Data. We're Filling It.

Most organisations have no shortage of data strategy or data tooling. What they consistently lack is someone to make it actually work.

Consultants Deliver Decks.

Strategy is presented. Roadmaps are agreed. Then the consultants leave — and nothing changes operationally.

Vendors Sell Licenses.

Platforms are deployed. Features are demonstrated. Then the implementation stalls and ROI never lands.

Nephos Delivers Outcomes.

We sit between both. We implement, operate, and stay. We make data capability real — not theoretical.

FINANCIAL SERVICES

Vanquis Bank

Head of Data Governance

UK retail bank · 1,000 employees · £500m revenue

60%cost reduction vs.
internal delivery**3**platforms
activatedregulatory audit
requirements met

THE CHALLENGE

- Need to enhance data quality and availability across the business
- Need for a consistent definition and understanding of data assets
- Audit requirement to demonstrate data management and quality capabilities relating to customer data

OUR APPROACH

- Technology selection and deployment for Data Catalogue, Business Glossary and Data Quality tools
- Use case design & execution with business level engagement to define business glossary terms
- Operational service providing ongoing support & administration of the platforms

THE OUTCOME

- 60% cost reduction compared to delivering the programme internally
- Platform providing common language across data terms used across multiple business units
- Data quality management used to meet audit requirements

The Race To Deploy GenAI Is Real

But the organisations that win aren't the fastest to deploy —

they're the first to make their data trustworthy enough to deploy on.

NEPHOS

o o o o

Thank You



www.nephos technologies.com



<https://www.linkedin.com/in/leebiggenden/>

