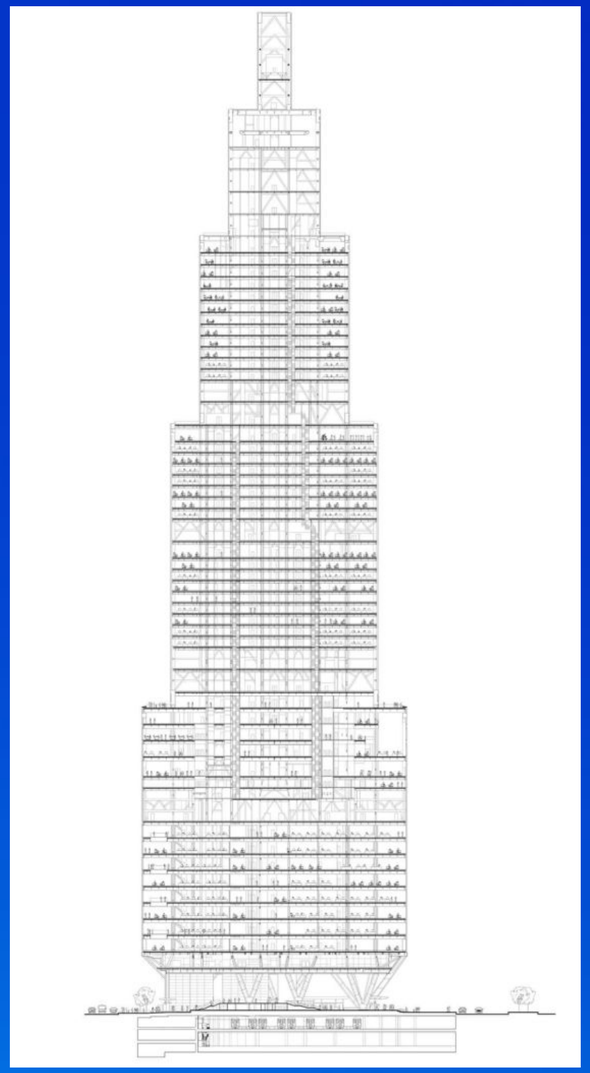
Beyond the Hype
How to Build the Foundation that Powers AI Success

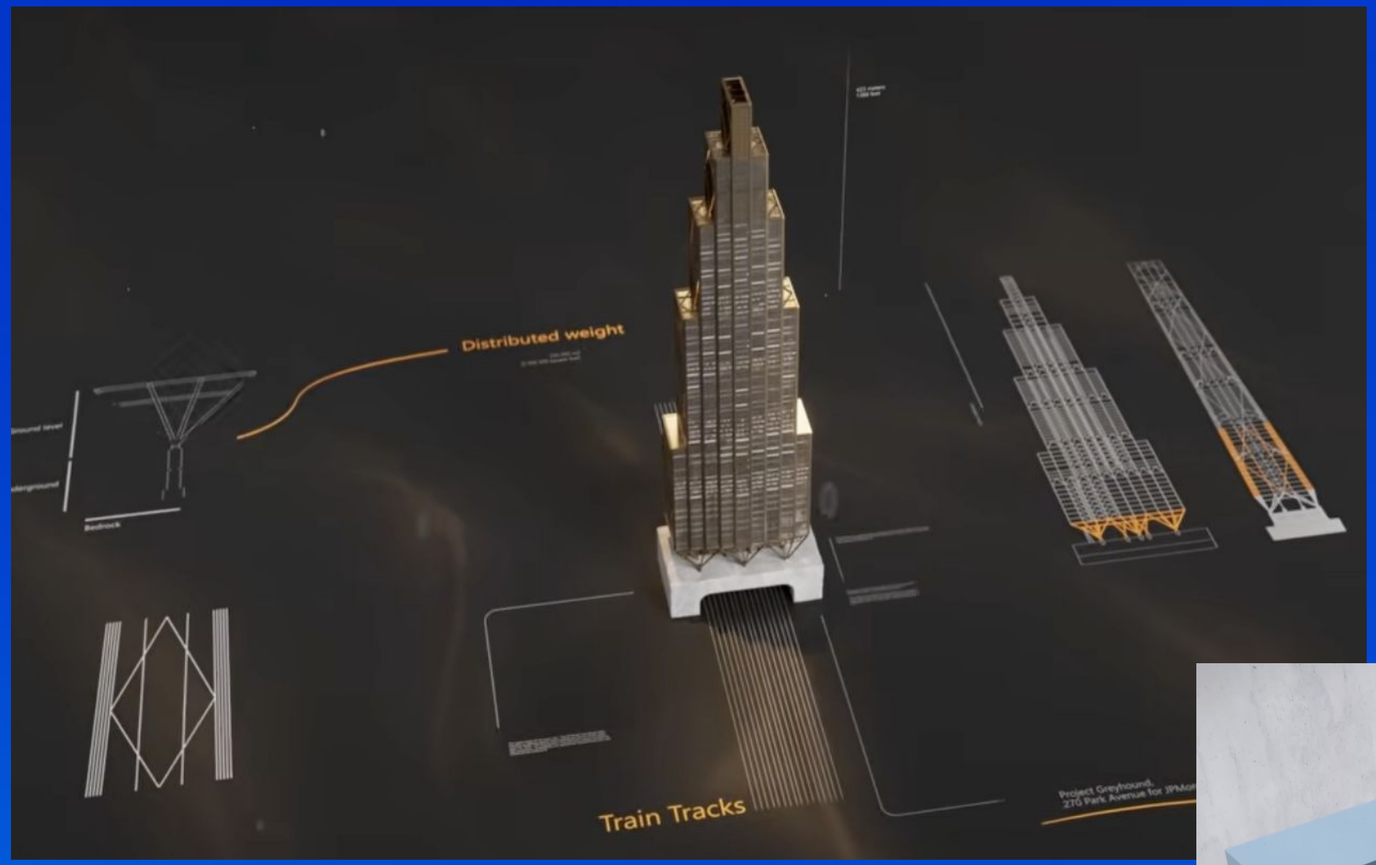
Joseph Sullivan
RVP, Solution Sales

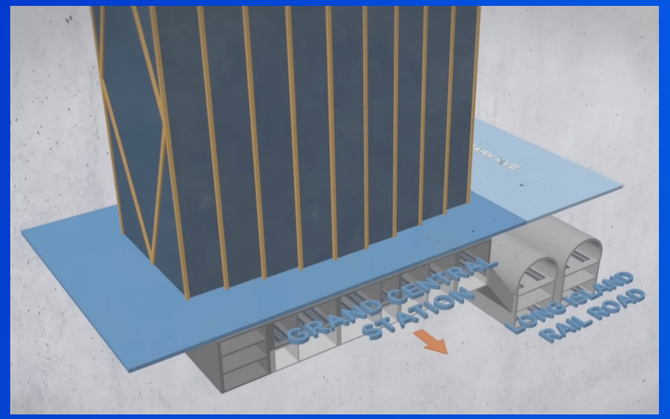


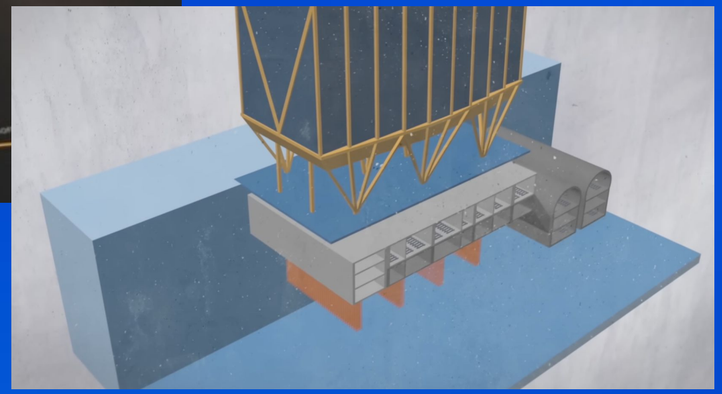


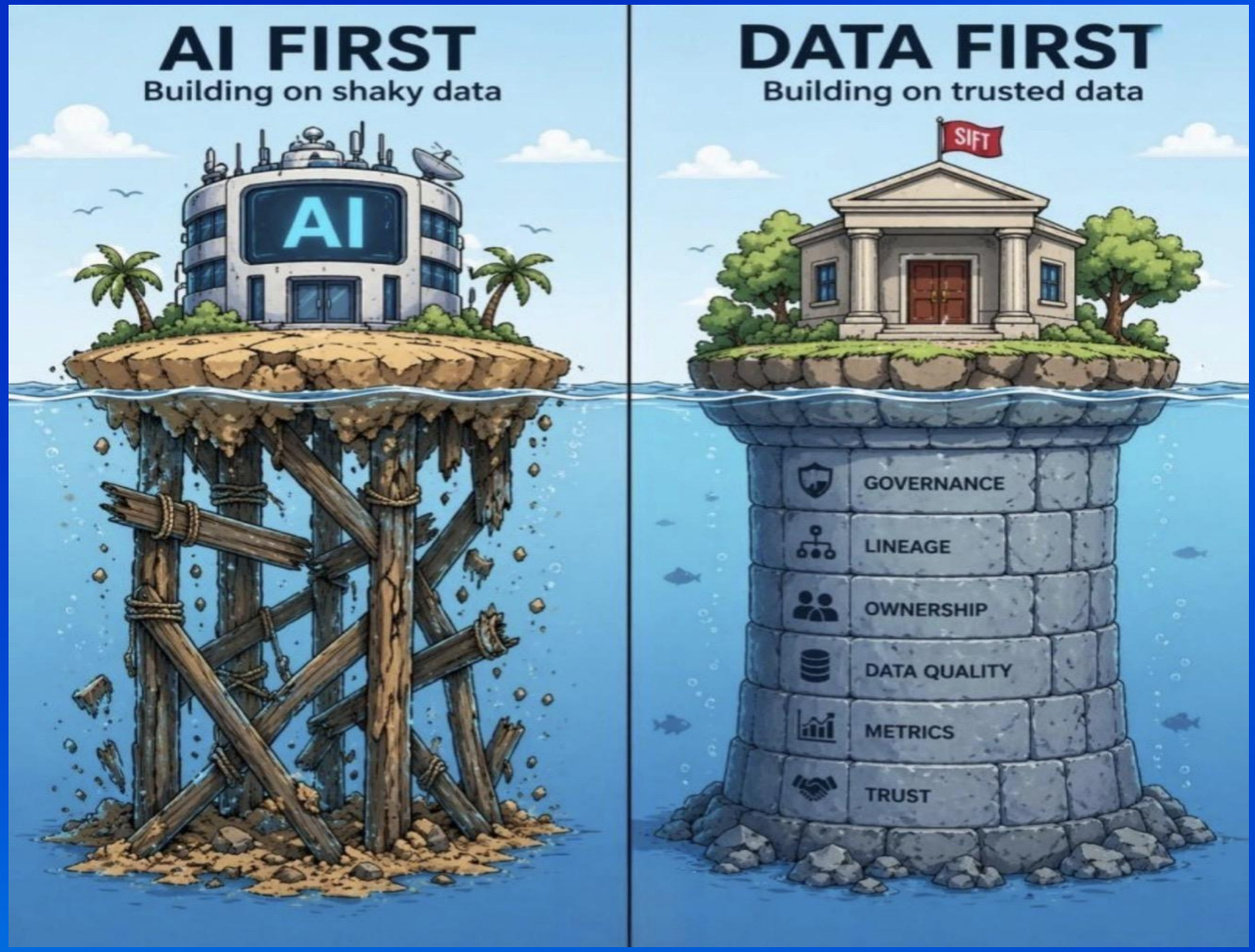





Distributed weight
Train Tracks


GRAND CENTRAL STATION
LONG ISLAND RAIL ROAD




AI FIRST
Building on shaky data
AI
DATA FIRST
Building on trusted data
SIFT
GOVERNANCE
LINEAGE
OWNERSHIP
DATA QUALITY
METRICS
TRUST

The AI Pilot Paradox

Most AI pilots succeed in the sandbox – and stall when they meet real enterprise data.

88%

of AI pilots never reach production

IDC / Lenovo, 2025

85%

of AI project failures caused by poor data quality

Gartner, 2025

7%

of enterprises say their data is ready for AI

HBR Analytic Services / Cloudera

Root Cause: Enterprise data is incomplete, inconsistent, and siloed. Sandbox AI uses curated data. Production AI meets reality.

The Data Reality

Data is accelerating in the Agentic Enterprise – and has never been more critical



You are at the center of this shift

From data custodians to **agentic data architects**

AI has moved from Answers to Actions

Question + Answers

Tasks + Actions

Multi-agents Orchestration

Robots

Predictions Suggestions

Automation

Super Agents

AGI

Most enterprises are deploying agents. Few have the data foundation to back them.

How many new customers did we add last quarter?



PRODUCT AGENT
287 Customers

MARKETING AGENT
1,104 Customers

FINANCE AGENT
512 Customers

SALES AGENT
340 Customers

Context

is the new currency



Unlocking Data is Critical to Context

Where data lives today:

Legacy on-prem databases and file systems

Enterprise apps: box, ORACLE, W

Hyperscalers: Microsoft, aws, Google Cloud

Salesforce Clouds:

Third-party sources: dun & bradstreet, zoominfo

- Lack of unified data
- Lack of real time context
- Untrusted data

Trapped Data, Rising Demands

Enterprises are sitting on a goldmine they can't access.
Data is everywhere – but fragmented and hard to trust.

Where data lives today:

Legacy on-prem databases and file systems

- **Enterprise apps:** SAP, Oracle, Workday
- **Hyperscalers:** Snowflake, Databricks, AWS, Google Cloud
- **Salesforce Clouds:**
- **Third-party sources:** dun & bradstreet, ZoomInfo

What that data must now power:

- Reporting & Analytics
- Application Activation
- AI Models & Predictions
- Business User Self-Service
- Autonomous Agents

Unlock → Trust → Activate

Three building blocks close the gap between trapped data and the outcomes your business demands.

Where data lives today:

Legacy on-prem databases and file systems

- Enterprise apps: SAP, Oracle, Workday
- Hyper-scalers: Snowflake, Databricks, AWS, Google Cloud
- Salesforce Clouds: Sales Cloud, Service Cloud, Marketing Cloud, Commerce Cloud
- Third-party sources: D&B, Zendesk

The three building blocks:

Unlock

Connect, Integrate, Understand

Trust

Validate, Protect, Unify

Activate

Decide, Orchestrate, Act

What that data must now power:

- Reporting & Analytics
- Application Activation
- AI Models & Predictions
- Business User Self-Service
- Autonomous Agents

Data requirements for the Agentic Enterprise

Every customer is trying to get their data right –
and their challenges consistently fall into three categories.

Unlock

Connect	Integrate	Understand
Streaming, Batch, Real-time, Zero Copy	API's, Apps, Legacy Systems	Semantics, Business Context and Intent

Trust

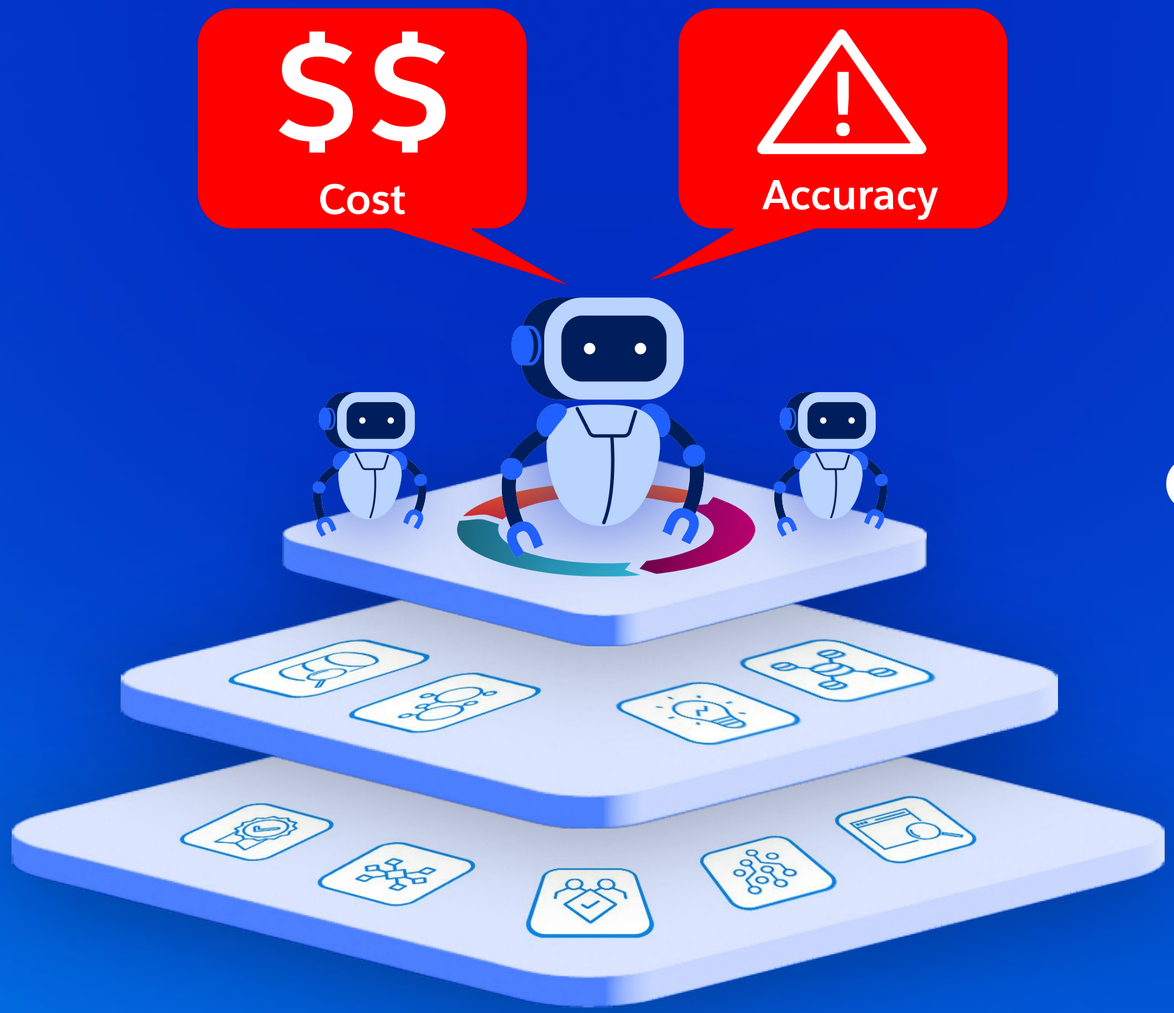
Validate	Protect	Unify
Discovery, Data Catalog, Data Lineage, Data Quality	Data Governance, Policies & Controls, Privacy Compliance	Master Data Management, Identity Resolution

Activate

Decide	Orchestrate	Act
Analytics, Predictions, Personalization, Segmentation	Workflows, Events, Agents, Processes	Agents, Analytics, Slack, Customer 360

Harmonization | Metadata | Ontology

What customers are saying



I understand the need for a data foundation, but I don’t have time to build it.

Can I just throw AI at it?

Accuracy

Does your AI know your business well enough to build the data foundation?

“Do I need a visa to visit the US for a two-week holiday?”

reddit

r/NZvisas • Posted by u/travelbro88 · 2019

▲ 312 upvotes

“Nah, Kiwis don't need a visa. Just book your flights and go.”

Popular, but unverified and out of date

U.S. Embassy in New Zealand

Official Guidance 2026 • nz.usembassy.gov

“New Zealand is in the Visa Waiver Program, so you don't need a visa. You must have an approved ESTA before you board.”

Verified, current and accountable

Cost

AI token usage and costs are out of control!

Uber

Burned through its entire 2026 AI budget in 4 months

Microsoft

Burned through its annual AI budget and is now cutting all Claude licenses by June 30

RBC

Token usage surged 500% in six months

Fortune 500

Racked up a $500 million Claude bill in a single month

Poor Model Selection



300 companies raised token cost concerns in earnings calls in April–May alone

Using AI for data management blows out token usage and costs



High Tech Co.

Identified spend of
$500k USD
on employees using AI to move files



Global Bank

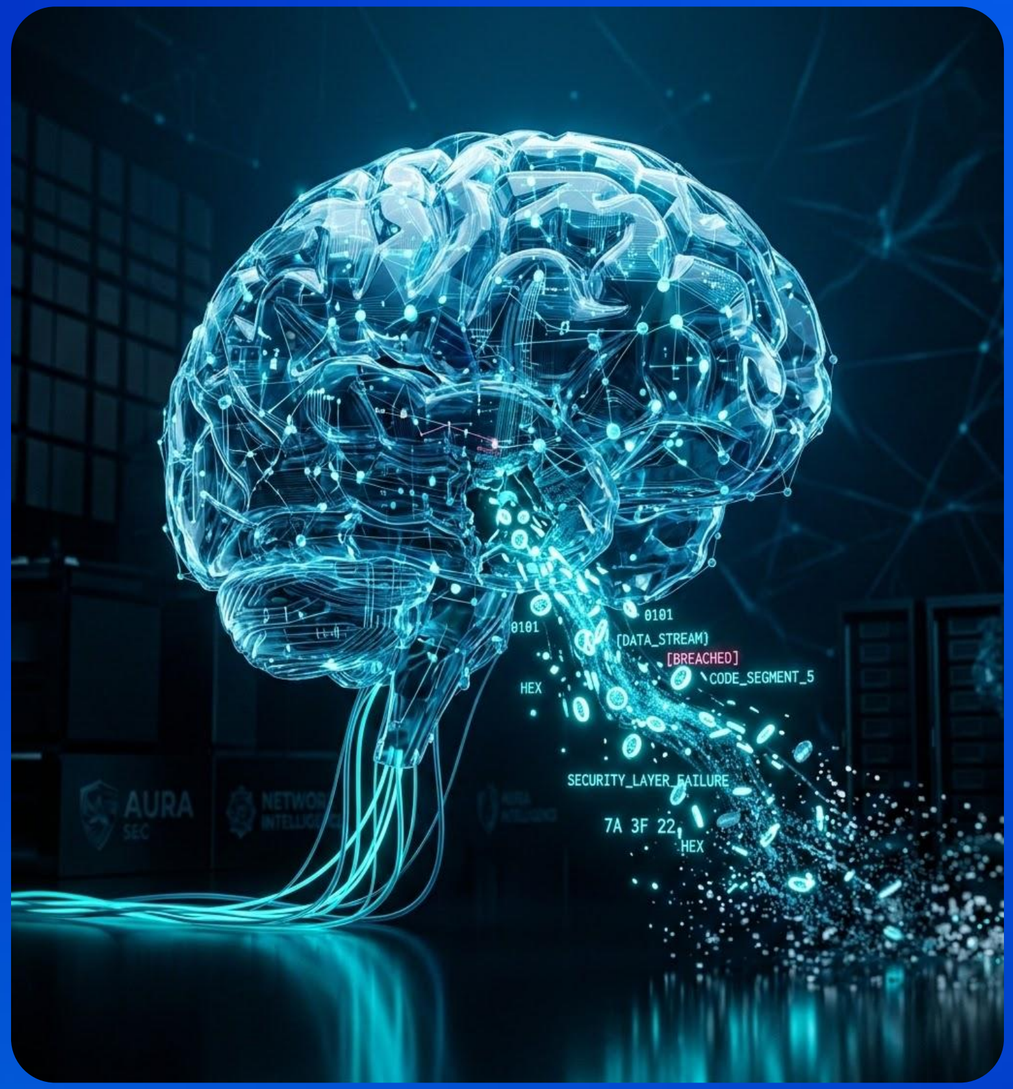
AI spend down by
95% ⬇ after
deployment of Data Governance

Data Management Token Bleed

when AI agents hunt through raw, ungoverned enterprise data to find what they need to build understanding of that data at runtime, burning repeat compute costs on every single query while still getting the wrong answer.

MCKNIGHT CONSULTING GROUP

Stop the Token Bleed: Benchmarking the Benefits of Governed Metadata for Enterprise AI



McKnight - Token Bleed Benchmarking Report

Experiment

NATURAL LANGUAGE PROMPT

"Show me all the assets needed to build a report that involves Government ID."

BENCHMARK TEST HARNESS

Enterprise Data Catalog (1.5M Objects - Snowflake, Oracle, S3)

↓

3 EVALUATED ACCESS METHODS using CLAUDE-OPUS-4.8

Direct DB Access | Context Stuffing | Governed Metadata

↓

LLM Response Token Consumption & Accuracy

Token Consumption (1.5M Scale)



The **Governed Metadata Layer** reduces token consumption to **2,112 tokens**, avoiding up to **29.8x token bleed** while maintaining 100% accuracy.

Source: McKnight Consulting Group Benchmark Report (1.5M Scale Tier)

McKnight - Token Bleed Benchmarking Report

Accuracy Metrics (1.5M Scale)

Governed Metadata Layer

F1: 1.000

100% Accuracy

2,264 Hits (TP)	0 Misses (FN)	0 Wrong Picks (FP)

Direct DB Access

F1: 0.592

59.2% Score

1,884 Hits (TP)	380 Misses (FN)	2,213 Wrong Picks (FP)

Context Stuffing

F1: 0.294

29.4% Score

1,448 Hits (TP)	816 Misses (FN)	6,122 Wrong Picks (FP)

Key Takeaways

1 Governed Metadata Ensures Absolute Accuracy

Without a governed data layer, system accuracy collapses. The Governed Metadata Layer achieved a perfect 1.00 F1 score with zero misses and zero wrong picks across all data scales.

FI accuracy of Context Stuffing and Direct DB access is equivalent to near-random performance.

2 Dramatic Reduction in Token Spend

Governed Metadata reduces LLM token consumption by up to 29.8× compared to Direct DB Access, bypassing costly context stuffing and output token bleed.

An agent with no governed path to data needs to build its own understanding of that data at runtime

3 Massive Financial ROI at Production Scale

On a workload of 10,000 agent calls per day, the governed route lowers annual cost from $2.9M to just $176K, delivering $2.7M in yearly savings.

Source: McKnight Consulting Group Benchmark Report (1.5M Scale Tier)


Informatica
from Salesforce
Thank you