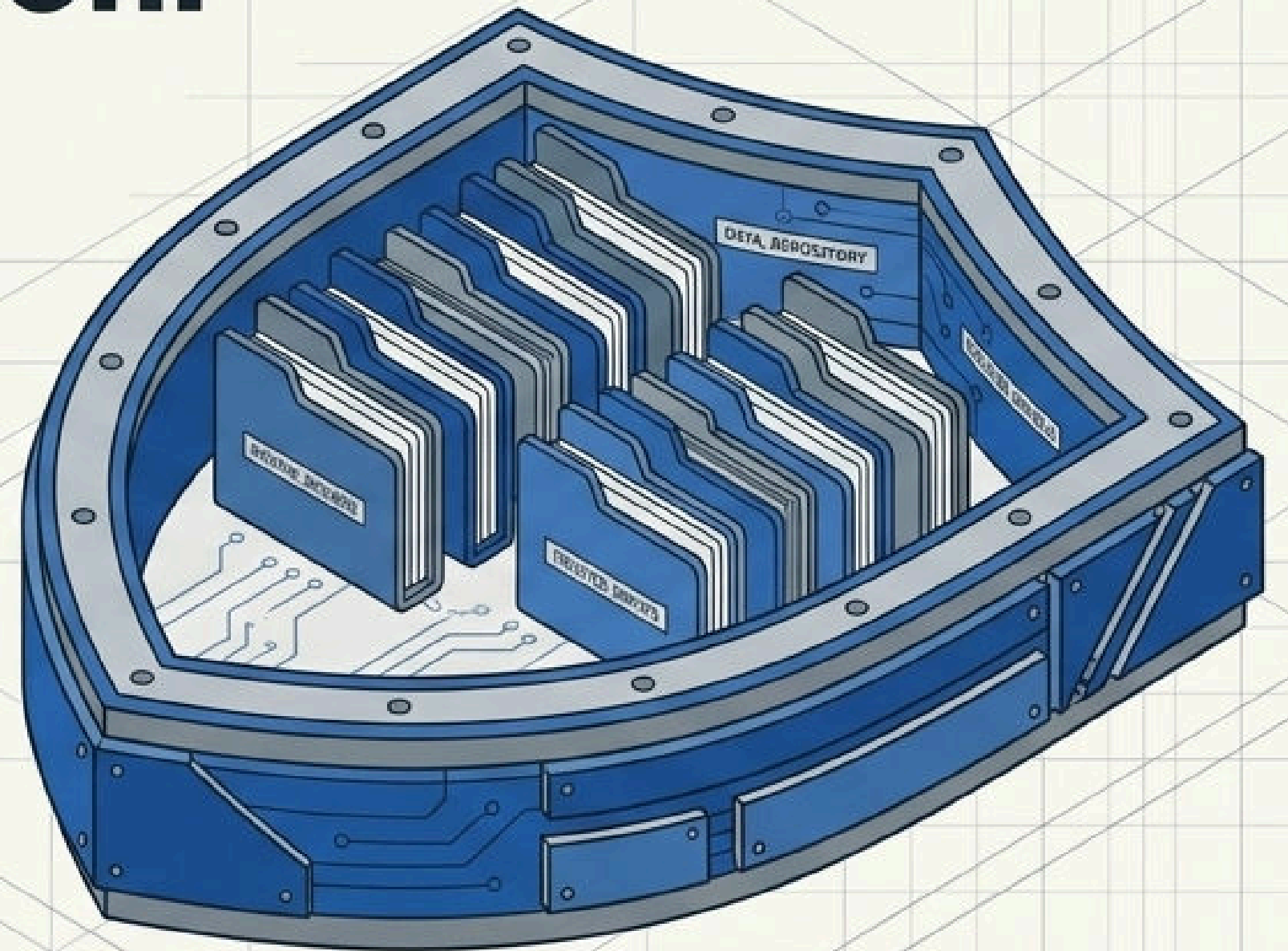


The Enterprise AI Agent Era demands a secure, governed data foundation.

A synthesis of the NIST AI Risk Management Framework and the Salesforce Architecture of Trust.

AI agents are only as reliable as the data they act upon.



About the Writers

- Global Top 20 Creator in AI Safety and Tech Ethics
- Corporate Trainer
- Creative Content Writer
- Top 20 Creator in Marketing and Sales on LinkedIn Singapore
- Growth-Driven Marketing Strategist
- Leading Brand Impact and Digital Transformation



Jean Ng



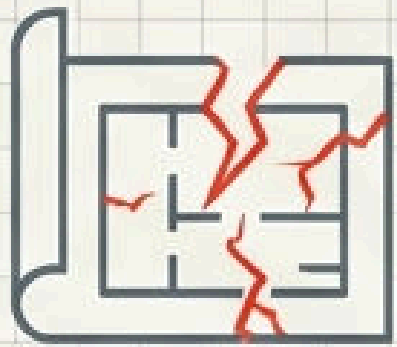
Gladys Ng

Follow us for more Marketing + AI content!



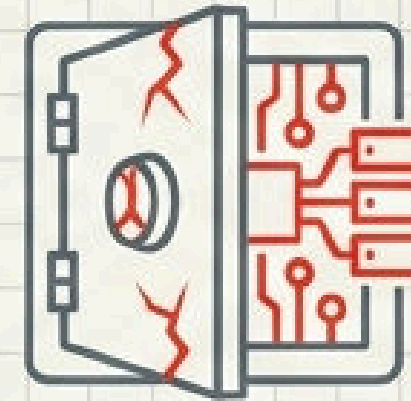
Repost to help your network.

Unchecked Generative AI introduces multidimensional enterprise vulnerabilities



Information Integrity

The unintentional or deliberate production of false, inaccurate, or misleading content, including systemic confabulations and disinformation.



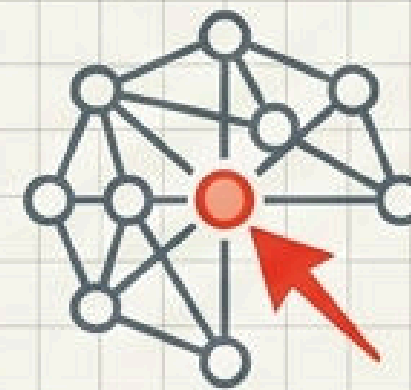
Data Privacy & IP

Unauthorized exposure of trade secrets, PII, and the eased replication of copyrighted or licensed content without authorization.



Societal Harm

The generation of harmful bias, homogenization, and obscene, degrading, or abusive content.



System Resilience

Increased attack surfaces for targeted cyberattacks and anomalous failure modes driven by adversarial manipulation.

The confidence gap is stalling enterprise AI deployment at scale.

48%

of IT leaders worry their underlying data foundation isn't ready for AI.

55%

lack confidence in their ability to implement AI with appropriate security guardrails.

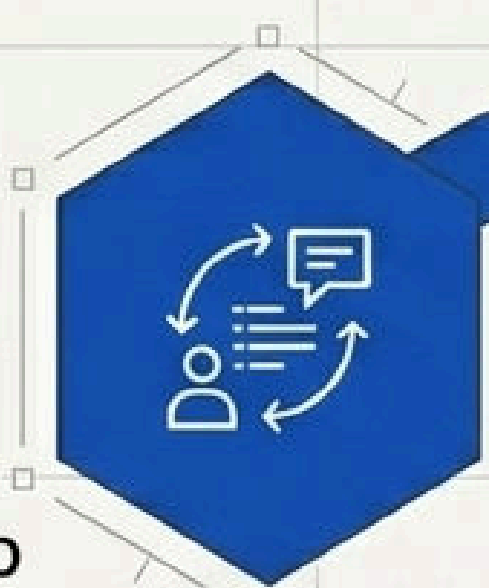
Impact Statement:

These concerns aren't just about preventing isolated errors; they're about ensuring compliance, maintaining data integrity, and building customer trust.

Rigorous evaluation requires continuous feedback loops and adversarial testing.

3. Gather Structured Feedback

Implement participatory engagement methods (user surveys, community forums) to monitor if outputs align with expectations and remain equitable.



1. Map the Context

Identify potential content provenance harms (deepfakes, tampered content) and ensure training data represents diverse in-context user populations.



2. Measure the Risk

Conduct adversarial role-playing and chaos testing (Red-Teaming) with both experts and the general public to identify unforeseen failure modes.



A resilient AI strategy begins with a unified, tightly governed data foundation.

Policy Enforcement:

High-scale metadata governance to manage access and enforce consistent policies across the enterprise.

Threat Detection:

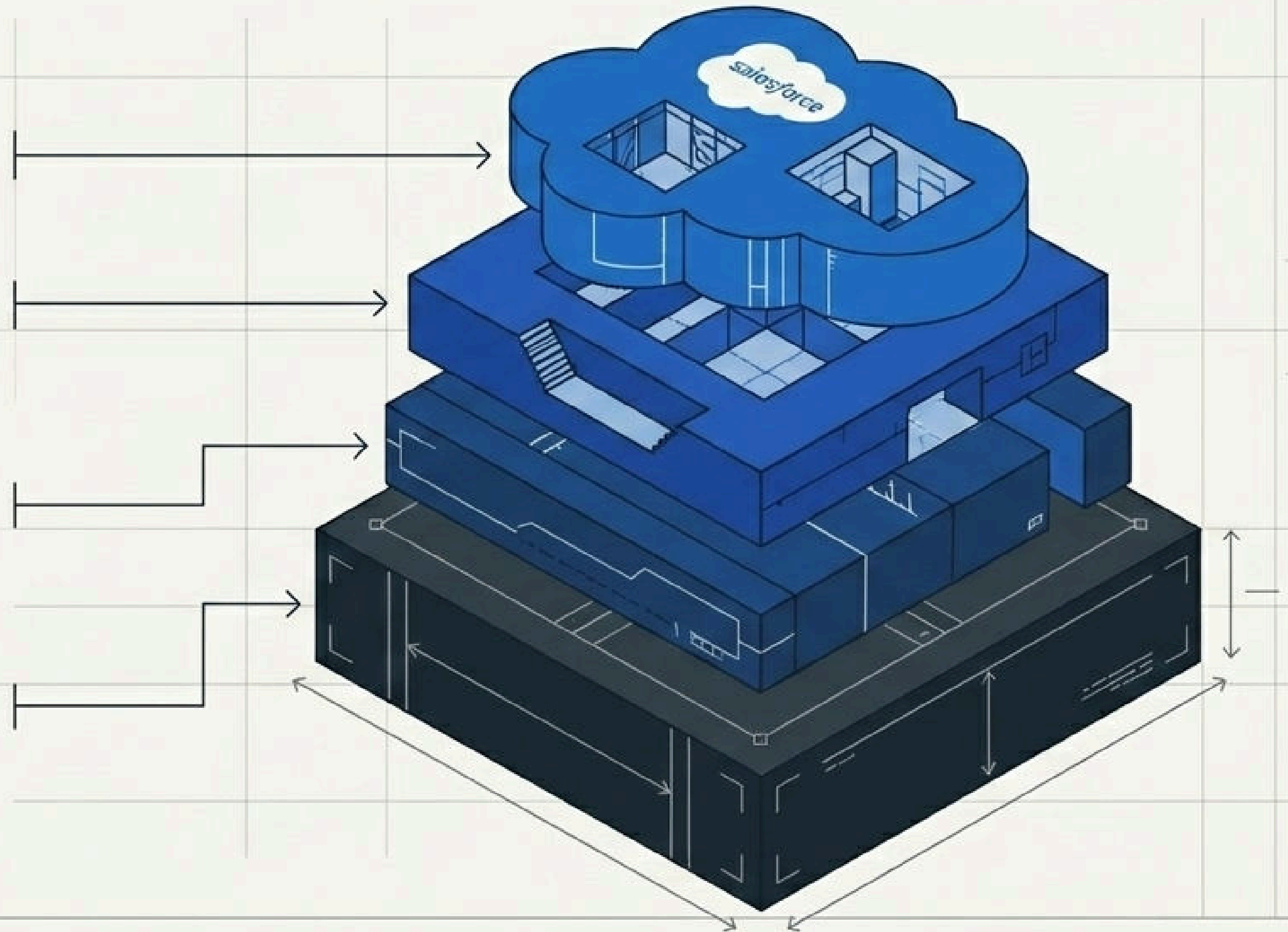
Deep visibility into user activity and proactive posture management with fast data classification.

Data Resilience:

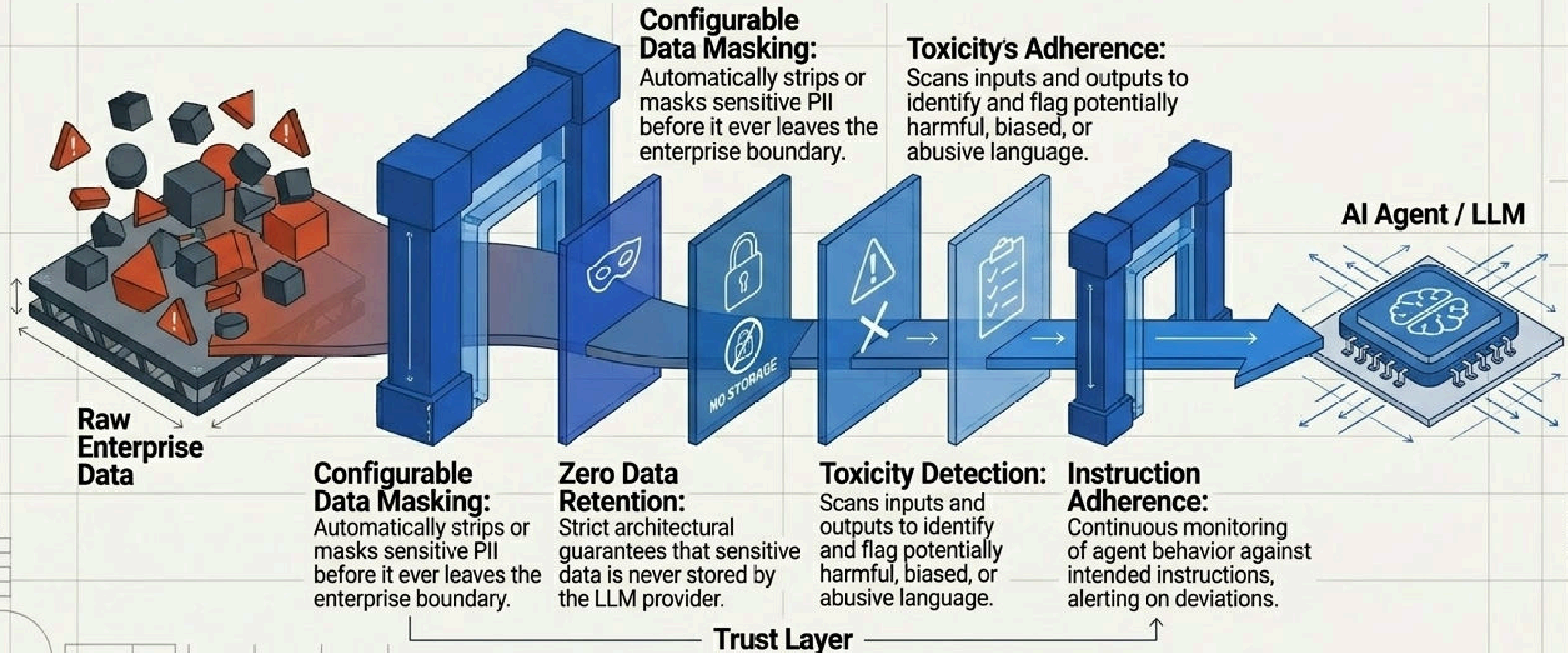
Automated backups and precise restoration to ensure 100% protection of critical metadata.

Unified View (Data Cloud):

Bringing together structured and unstructured data to deliver a trusted, actionable view of the customer.



The Trust Layer acts as an impenetrable governance filter between your data and the LLM.



The strategic blueprint for deploying trustworthy Agentic AI.



1. Mitigate and Mask at the Source

Never expose raw data. Utilize automated masking and zero-retention policies to ensure models cannot learn from or leak proprietary enterprise data. (NIST MANAGE + Salesforce Trust Layer)



2. Implement Real-Time Content Filters

Deploy rule-based and machine learning models to analyze generated content for toxicity, bias, and instruction adherence before it reaches the end user. (NIST MANAGE 3.2)



3. Sustain Value through Continuous Auditing

Establish ongoing, structured feedback mechanisms and utilize explainable AI (XAI) techniques to constantly trace the lineage, authenticity, and impact of AI-generated actions. (NIST MANAGE 2.2)

Sources:

- 1) NIST, Artificial Intelligence Risk Management Framework: Generative AI Profile
- 2) McKinsey, State of AI trust in 2026: Shifting to the agentic era
- 3) Salesforce, Why Trust, Security, and Governance are Non-Negotiable