

The Hidden Cost of Unstructured Data: Why AI Model Training Needs a New Approach

1

The Critical Role of Unstructured Data

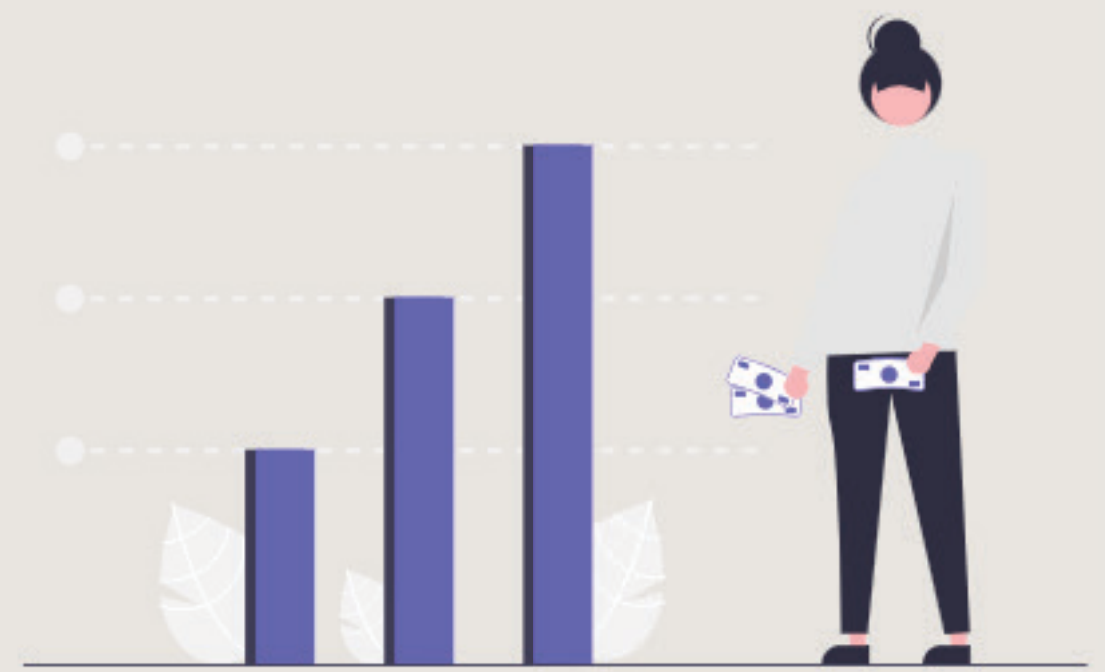
Unstructured data, accounting for 80–90% of enterprise data and expanding at a rapid 55–65% annually, is essential for powering generative AI, LLM fine-tuning, and Retrieval-Augmented Generation (RAG). Yet, only 18% of organizations harness it effectively, highlighting its untapped potential as a cornerstone for AI innovation.



2

Challenges of Training AI Models with Unstructured Data

Inconsistent data hampers AI accuracy, drives up costs, and complicates compliance, while sensitive data increases security risks. Inefficient structures introduce bias, and storage demands strain resources, highlighting the need for smarter unstructured data management.



3

The Shift to Decentralized Data Management

Decentralized data management distributes control across teams, enabling data refinement at the source through local tagging, filtering, and organization. This ensures accuracy, security, and compliance while empowering AI to deliver precise outcomes and fostering accountability for better decision-making.



4

6 Steps to Decentralized Data Management

1. Establish data governance with clear policies, standards, and automation for secure compliance.
2. Build a data fabric for real-time access using metadata management and virtualization.
3. Use federated access with secure credentials, multi-factor authentication, and adaptive controls.



5

4. Leverage AI-driven classification to enhance governance and reduce errors.
5. Integrate on-premises and cloud systems with hybrid orchestration for optimized resource usage.
6. Choose self-service unstructured data management software for quick deployment, security, and scalability.

