Use Case

AI in Action: Building AI-Ready Data at Scale with Zubin

In today's AI-powered economy, data isn't just an input—it's the engine behind every decision, prediction, and innovation. But here's the critical truth most enterprises overlook: AI is only as intelligent and secure as the data it's fed.

And yet, over 80% of enterprise data remains unstructured, sitting in silos, uncatalogued, unclassified. According to Gartner, by 2026, 75% of AI projects will underperform or fail due to poor data quality, bias, and lack of governance. When AI models ingest ungoverned data, the result is not innovation it's risk.

The consequences?

- Biased or inaccurate outputs that derail business decisions
- Security breaches from hidden PII/PHI embedded in training datasets
- Compliance violations due to cross-border data movement or misuse
- Reputational damage from AI-generated hallucinations and trust failures

Here's the core issue:

You can't build intelligent AI if you don't know what your data contains, where it resides, or whether it should even be used.

This is where Zubin changes the game.

Zubin, Data Dynamics' AI-powered self-service data management software, brings intelligence, security, and control to the data that feeds your AI, making it AI-ready by design.

Three Core Challenges Slowing AI Readiness

The Unstructured Data Blind Spot

Data lives across object stores, NAS shares, SharePoint sites, and cloud archives unstructured,uncatalogued, and invisible to AI pipelines. Without automated discovery, organizations unknowingly feed sensitive or irrelevant data into their models.

The Risk and Compliance Minefield

PII, PHI, IP, and jurisdiction-bound data often slip through into training sets, triggering legal violations. Without intelligent classification & localization enforcement, AI readiness becomes a compliance liability.

The Explainability Gap

Most enterprises cannot trace what data went into a model, how it was processed, or whether it met internal policies, creating a black box that regulators and executives alike can't trust.

This Is Where Zubin Helps

Zubin addresses the entire AI data lifecycle—from discovery and classification to governance, optimization, and compliance—giving enterprises the ability to feed AI with clean, secure, and contextually intelligent data.

Zubin empowers enterprises to

- Discover and classify unstructured data across hybrid environments without manual intervention
- Detect sensitive content like PII/PHI, contracts, IP, and source code using AI-powered analytics
- Enforce sovereignty and localization mandates with built-in geo-fencing policies
- Control access with RBAC and policy-based governance to ensure only the right users and systems feed your AI
- Optimize and eliminate ROT data to reduce compute waste and improve training accuracy
- Build traceability into every dataset used, supporting explainability and regulatory compliance

Zubin isn't just a data tool—it's your AI data control plane.

Click here for a demo

What's In It for You?

Accelerate AI Data Prep with Intelligent Discovery

Zubin uses content and metadata analytics to scan petabytes of unstructured data across cloud and on-prem environments—automatically classifying what's usable, what's risky, and what should be excluded from AI workflows.

Ensure Compliance with Built-In Sovereignty and Risk Controls

With preconfigured policy packs aligned to laws like GDPR, DPDP, HIPAA, and PDPL, Zubin automatically prevents cross-border violations, mitigates exposure, and applies redaction or quarantine to non-compliant data.

Govern Model Inputs with Role-Based Access and Actionability

Instead of centralized bottlenecks, Zubin enables federated self-service governance. Data owners and application teams can apply controls, trace usage, and govern what enters the AI pipeline securely and at scale.

Enable Transparent, Trustworthy AI Through Lineage and Auditability

Zubin embeds full audit trails and lineage into every dataset prepared for AI, including classification context, access logs, policy history, and movement traces. This supports internal governance, model transparency, and regulatory defensibility.

Reduce Waste and Risk with Storage Lifecycle Management

Zubin identifies ROT data and archives, deletes, or re-tiers it, reducing storage costs, minimizing training noise, and focusing AI compute on what matters.

A Closer Look

Zubin's Core Capabilities Powering AI Readiness

Zubin offers a comprehensive suite of intelligent, automated capabilities that transform raw, unstructured data into curated, trusted, and AI-ready datasets. These capabilities not only enhance the accuracy and compliance of your AI models but also reduce operational overhead by automating the most complex elements of unstructured data management.

AI/ML-Based Discovery and Content Classification

At the heart of AI-readiness lies the ability to accurately detect & understand what data exists in your environment. Zubin uses a dual-engine approach—combining metadata analytics with deep AI/ML-powered content inspection to detect:

- PII/PHI (names, IDs, financial records, medical terms)
- Intellectual property and source code
- Contractual and regulated documents
- Unstructured formats like PDFs, chat logs, scanned images, email threads

Zubin builds context-aware classification tags, enriched with domain specific insights to ensure the right data is used, while risky or irrelevant content is flagged or excluded from AI pipelines.

Role-Based Access Control (RBAC) and Policy Enforcement

AI requires not just access, but the right access. Zubin's RBAC framework ensures that only the right data owners, data scientists, and application teams can view or act upon datasets destined for AI processing.
Key capabilities include:

- **Policy-Based Permissions**
  Automates access decisions based on role, sensitivity level, location.
- **Open Share Reporting**
  Identifies misconfigured files or folders that may expose AI models to unauthorized or shadow data.
- **File Re-Permissioning**
  Automatically updates permissions in bulk to minimize overexposure.

This level of control is essential to prevent AI models from ingesting unauthorized or non-compliant data.

Data Usage Monitoring, Lineage & Traceability

To build trusted AI, you need to be able to prove what data was used, who accessed it,and how it was transformed. Zubin builds end-to-end lineage and traceability into every dataset:

- **Usage Analytics**
  Tracks data access patterns and anomalies in AI workflows.
- **Lineage Mapping**
  Captures source, versioning, transformation, and user-level actions.
- **Audit Trails**
  Stores logs for regulators, internal audits, and model validation.

This capability supports AI explainability, improves governance, and enables defensible regulatory compliance.

Risk Analysis and Automated Remediation

Zubin enables real-time, intelligent risk assessment of AI training data by evaluating:

- Exposure of sensitive data within AI-accessible environments
- Inadequate permission models
- Inherited regulatory obligations (e.g., GDPR, HIPAA, DPDP)

Once risks are identified, Zubin automatically initiates remediation using defined workflows such as:

- Redacting sensitive fields (e.g., patient names or employee IDs)
- Archiving or quarantining risky files
- Re-permissioning shared directories
- Flagging high-risk assets for review

This level of control is essential to prevent AI models from ingesting unauthorized or non-compliant data.

Data Sovereignty and Privacy Protection

With the surge in Sovereign AI and data localization mandates, AI-readiness now includes jurisdictional compliance. Zubin allows you to:

- Enforce geo-fencing of data used for AI training based on the origin country
- Prevent cross-border access of data covered by GDPR, PDPL, or India's DPDP Act
- Detect and restrict PII/PHI leakage into generative AI and LLM pipelines

By combining location-aware controls with intelligent classification, Zubin ensures your enterprise meets AI regulatory requirements without operational delays.

Storage Lifecycle Optimization for AI Readiness

Not all data should be fed into AI. Zubin identifies Redundant, Obsolete, or Trivial (ROT) data and applies lifecycle rules to:

- Archive cold or legacy data
- Eliminate irrelevant files from AI datasets
- Tier data across storage based on usage and relevance

This ensures that AI pipelines are powered by lean, high-quality, and current data, reducing compute waste and increasing model accuracy.

AI Readiness, Without the Guesswork

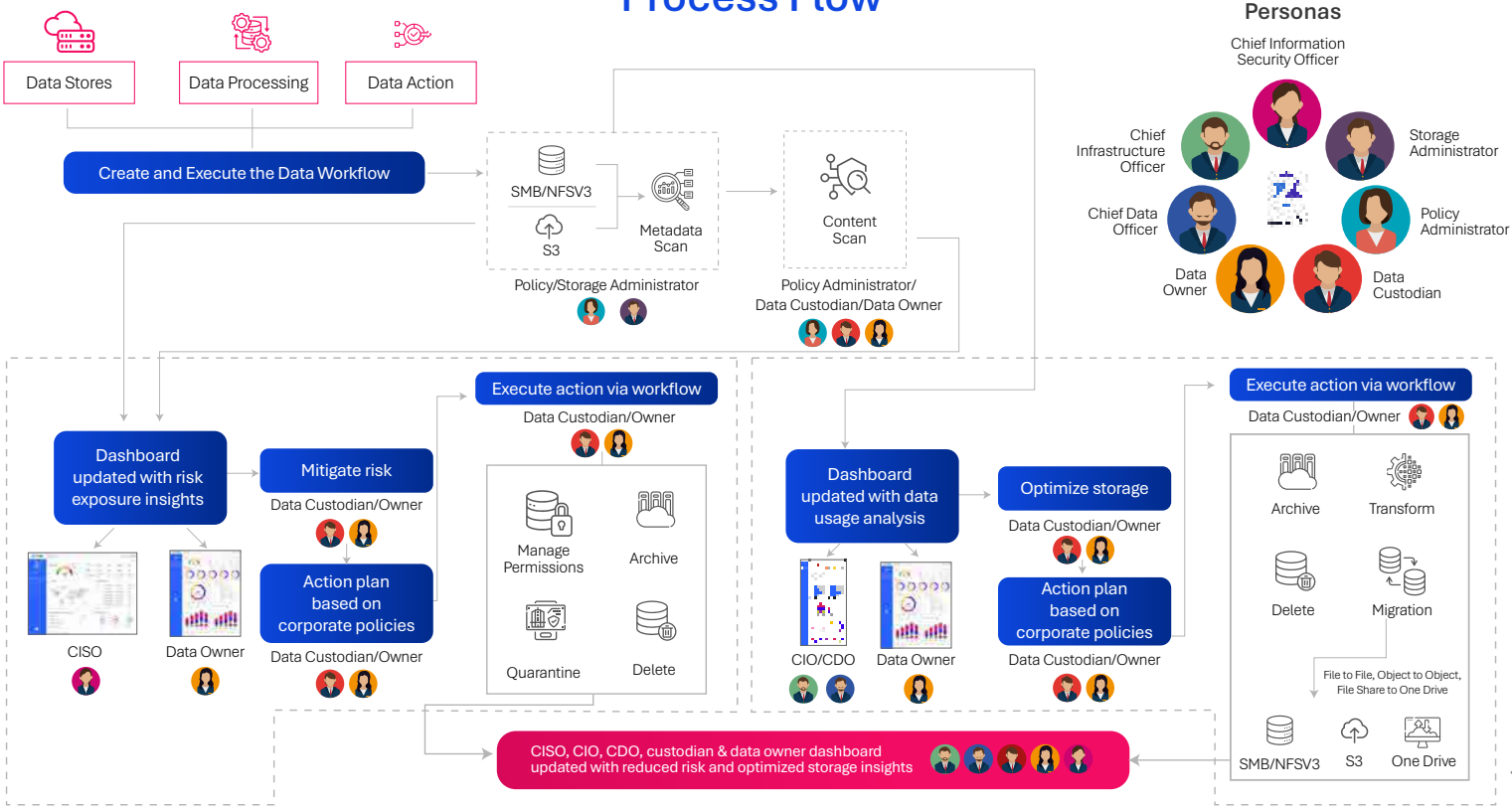
Feeding AI the wrong data doesn't just lead to poor decisions—it creates an enterprise-wide risk surface.

With Zubin, enterprises don't just prepare data for AI—they govern it, trace it, and trust it.
Because in 2025, AI readiness isn't a checkbox—it's a capability.
And Zubin makes sure your data is ready for what's next.

A Closer Look at What Makes Zubin Stand Out

- Enterprise Risk Posture & Infra Usage BI & Reporting
- Data Risk & Usage by Team & Data Ownership
- Personalized Data Policy Creation & Workflow
- Metadata Analytics
- Content Analytics
- Data Classification
- Statistical Sampling
- Data Migration
- Data Minimization
- Data Archival
- Retention Compliance
- Data Transformation
- Data Tiering & Placement
- Cloud Integrated Risk Controls
- Access Control & File Re-permissioning
- Data Sharing
- Risk Exposure Insights
- Data Containment & Isolation
- Real-time Notifications and Reporting

Process Flow



Your next chapter of success awaits; let's write it together with Zubin. **Click here for a demo**