ARTIFISCAL COMPANION OPERATIONAL PLAYBOOK

THE ARTIFISCAL DATA GOVERNANCE FRAMEWORK



Building the Clean Data Perimeter for Enterprise AI — Companion to The Modernization Manifesto

Jonathan Anderson | Managing Principal - Enterprise AI Strategy, Artifiscal

Executive Overview: Garbage In, Liability Out

The Modernization Manifesto establishes a clear boundary for enterprise leadership: artificial intelligence must be governed as an operating discipline, not deployed as an ad-hoc software experiment. Yet even the strongest AI charter will collapse if the underlying data is fragmented, outdated, or unverified.

An AI model is only as trustworthy as the information perimeter that feeds it. When an organization connects generative models or autonomous agents to messy legacy databases, duplicate spreadsheets, and unvetted document repositories, it does not automate efficiency. It automates confusion at machine velocity.

In *The Modernization Manifesto*, Green-Light Condition #2 states that no workflow is ready for AI deployment until the data perimeter is clean. This companion framework provides CIOs, Chief Data Officers, and operational leaders with the tactical blueprint to build and maintain that perimeter.

Two Operating Principles Drive This Framework:

1. **Data Rigor Protects Return on Token (ROT):** Every time an AI system retrieves conflicting records, processes redundant files, or requires human rework to fix a hallucinated output, the enterprise burns compute budget without generating margin. Clean data is not an IT hygiene chore. It is a direct lever for financial performance.
2. **Minimum Viable Governance:** Most enterprise data initiatives stall because teams attempt to catalog every file in the company before delivering value. Owner's Rigor favors **Minimum Viable Governance**: identifying and locking down the 20% of Critical Data Elements (CDEs) that power 80% of revenue, risk, and customer experience first, then expanding outward.

Part I: Data Classification & AI-Eligibility Tiers

Traditional data security labels (Public, Internal, Confidential) were built for static file storage and human email. They break down in an environment where employees paste proprietary text into browser-based tools and autonomous agents query live databases via APIs.

To contain Shadow AI and enforce *The Law of Harm Mitigation*, enterprises must add an **AI-Eligibility Layer** to every data asset before it can be ingested into a model, RAG (Retrieval-Augmented Generation) pipeline, or third-party platform.

The 4-Tier AI-Eligibility Classification Matrix

AI-Eligibility Tier	Data Profile	Permitted AI Usage	Representative Examples	Mandatory Technical Controls
Tier A: Open	Publicly released, non-proprietary information.	Approved for commercial external LLMs and public APIs.	Published marketing collateral, press releases, public product documentation, SEC filings.	Standard acceptable-use logging; no special masking required.
Tier B: Internal / Private AI Only	Proprietary operational data without regulated PII or trade secrets.	Restricted strictly to enterprise-licensed, zero-retention private AI environments.	Internal SOPs, anonymized operational metrics, non-sensitive project plans, training manuals.	Enterprise tenant isolation; contractual zero-training guarantees; SSO authentication.
Tier C: Restricted / Masked Only	Sensitive commercial, financial, or customer data carrying legal or competitive risk.	Permitted only inside private models **after** automated PII/entity redaction or tokenization.	Customer records, claims histories, vendor pricing contracts, unannounced financial forecasts.	Automated DLP (Data Loss Prevention) masking at the prompt/API layer; audit logging; Tier 2 HITL review.
Tier D: Prohibited / Air-Gapped	Crown-jewel intellectual property, regulated health/payment data, or privileged legal records.	**Hard prohibition.** Excluded from all generative AI prompts, training sets, and agentic workflows.	Source code for proprietary algorithms, M&A term sheets, executive litigation files, unencrypted credentials.	Hard architectural blocks at the API gateway; zero indexing by internal AI search tools.

Part II: Data Quality, Lineage & Grounding

As established in *The Law of Transparency*, artificial intelligence does not lie out of malice. It hallucinates and exhibits bias when it is forced to generate a plausible answer across an information gap. Grounded AI requires verified, consistent data.

2.1 The Six Dimensions of Enterprise Data Quality

Before any dataset is cleared for an AI workflow or API integration, its steward must verify it against six measurable quality standards:

1. **Accuracy:** Does the record reflect verified real-world truth (e.g., current policy limits, validated customer addresses)?
2. **Completeness:** Are mandatory fields populated, or will missing values force an algorithm to guess?
3. **Consistency:** Does the same entity (customer, vendor, SKU) carry identical values across CRM, ERP, and billing platforms?
4. **Timeliness:** Is the data refreshed on a cadence fast enough to support the decision being made, or is the model reasoning over stale exports?
5. **Validity:** Do values conform to strict structural formats, ranges, and business rules?
6. **Uniqueness:** Have duplicate records been eliminated so models do not double-count revenue, risk, or customer interactions?

2.2 Master Data Management (MDM) & Source-of-Truth Lineage

One of the fastest ways to trigger enterprise hallucinations is pointing an internal AI assistant at an ungoverned shared drive containing six conflicting versions of a policy manual or pricing sheet.

- **Golden Records Over Copies:** Establish a single authoritative "System of Record" for core entities (Customers, Products, Financials, Employees). AI pipelines must pull exclusively from these governed golden records, never from local spreadsheets or unverified departmental folders.
- **End-to-End Lineage:** Every output generated by a Tier 2 (Augmented) AI system must be traceable back to its exact source document, database table, and timestamp. If a human reviewer cannot verify where a number or clause originated in under 60 seconds, the pipeline fails the transparency standard.

Part III: Access Control, Privacy & Vendor Boundaries

3.1 Dual-Identity Access Control (Humans and Agents)

In legacy IT environments, permissions were assigned to human employees. In a modernized enterprise, AI agents and automated workflows also act as system users. Without strict governance, an internal AI assistant indexed across company servers can accidentally surface executive compensation or confidential HR files to an entry-level employee.

Organizations must enforce **Permission Inheritance and Least Privilege**:

- **Zero Permission Escalation:** An AI agent or RAG search tool must inherit the exact Role-Based Access Control (RBAC) permissions of the human user submitting the prompt. If the human cannot open a file directly, the AI cannot retrieve or summarize it on their behalf.
- **Scoped Service Accounts:** Automated Tier 1 workflows must run on dedicated, narrowly scoped API tokens with read/write access limited strictly to the single table or application required for that task.

3.2 Data Retention, Sunsetting, and Vendor Contracts

Data that sits indefinitely in cloud storage without a business purpose is not an asset. It is unpriced breach liability and unnecessary compute overhead.

- **Automated Sunsetting:** Assign explicit retention schedules to every Critical Data Element. Archive or purge expired records automatically so models never train on or retrieve obsolete business rules.
- **Non-Negotiable Vendor Clauses:** Mirroring the *Lifecycle Transparency Audit* in *The Modernization Manifesto*, every third-party AI or SaaS contract must include three hard legal boundaries:
  1. **Zero Data Retention (ZDR)** on API prompts and completions.
  2. **Explicit Prohibition on Model Training** using your enterprise inputs, outputs, or metadata.
  3. **Certified Deletion Guarantees** verifying that all company data and fine-tuned weights are permanently purged within 30 days of contract termination.

Part IV: Data Stewardship & Accountability

Software tools do not maintain data hygiene; people do. To operationalize *Relational Governance* (Safety, Benefit, and Stewardship) across the data lifecycle, every Critical Data Element must have a named human owner.

The Three-Role Data Accountability Model

Governance Role	Typical Title / Function	Core Responsibility	Governance Alignment
Executive Data Owner	VP of Finance, COO, Chief Underwriting Officer, General Counsel	Holds final business accountability for a data domain; approves Tier C/D classifications and external vendor sharing.	**Safety & Accountability**
Domain Data Steward	Senior Ops Manager, Lead Business Analyst, Controller	Defines business rules, monitors the six quality dimensions, resolves conflicting definitions, and approves RAG knowledge bases.	**Benefit & Operational Fit**
Technical Data Custodian	Data Engineer, Cloud Architect, IT Security Lead	Engineers the pipelines, enforces RBAC/DLP masking, maintains lineage logs, and executes automated retention schedules.	**Stewardship & Architecture**

Plugging Into the Monthly AI Governance Council

Do not create a separate bureaucracy for data governance. Embed Data Owners and Lead Stewards directly into the monthly **AI Governance Council** established in *The Modernization Manifesto*:

- No new AI use case can pass Phase 1 intake without a signed Data Readiness sign-off from the relevant Domain Data Steward.
- Quarterly drift audits review both model accuracy and underlying data defect rates side by side, ensuring teams fix root-cause data problems rather than repeatedly re-prompting broken workflows.

Part V: The 90-Day Data Hygiene Sprint & Board KPIs

Enterprises do not need a multi-year transformation program to build a clean data perimeter. Using **Minimum Viable Governance**, leadership can establish foundational data control in 90 days, running directly alongside Phase 1 and Phase 2 of the *180-Day "Build + Run" Roadmap*.

The 90-Day Execution Sprint

Days 1 to 30: Audit & Catalog Critical Data Elements (CDEs)

- Identify the top 2 to 3 target workflows slated for AI or API automation.
- Map the exact source systems, tables, and document repositories feeding those workflows.
- Assign an Executive Data Owner, Domain Steward, and Technical Custodian to each domain.

Days 31 to 60: Classify, Clean & Lock Permissions

- Tag all in-scope assets using the **4-Tier AI-Eligibility Matrix** (Tiers A through D).
- Deduplicate records, archive stale document versions, and establish a single Golden Record for each core entity.
- Enforce RBAC permission inheritance and deploy automated DLP/PII masking at the API gateway.

Days 61 to 90: Automate Quality Gates & Launch Monitoring

- Implement automated validation rules that flag missing fields, schema breaks, or anomalous values before data enters an AI pipeline.
- Activate lineage tracking and begin reporting monthly Data Governance KPIs to the AI Governance Council.

Board-Ready Data Governance KPIs

Metric	What It Measures	Target Benchmark	Executive Value
AI-Eligibility Classification Rate	% of active repositories and databases tagged under Tiers A to D.	**100% of in-scope AI systems**	Eliminates blind spots and prevents accidental leakage of Tier D assets.
Critical Data Lineage Coverage	% of Tier 2 AI outputs traceable to a verified Golden Record.	**100%**	Enforces *The Law of Transparency* and ensures audit readiness.

Metric	What It Measures	Target Benchmark	Executive Value
Data Defect Rate	% of records failing accuracy, completeness, or uniqueness checks prior to ingestion.	**< 1.0% on Critical Data Elements**	Stops hallucinations and downstream operational errors at the source.
Mean Time to Remediate (MTTR)	Average hours required for a Steward/Custodian to resolve a flagged data anomaly.	**< 24 hours for Tier 1/2 pipelines**	Measures active operational rigor rather than passive reporting.
Token Waste Ratio	% of departmental AI token spend tied to reprocessing, duplicate queries, or bad-data rework.	**< 5.0% of quarterly token budget**	Connects data hygiene directly to CFO margin and **Return on Token (ROT)**.

Companion Diagnostic: The Data Readiness Checklist

Before approving any AI vendor, internal RAG deployment, or automated workflow, score your target data domain using this 10-point diagnostic. Score **1 point** for every "Yes" and **0 points** for every "No" or "Unsure."

#	Data Readiness Diagnostic	Response
1	**Critical Data Identification:** Have you explicitly identified the exact source tables and documents required to power this workflow?	☐ Yes ☐ No
2	**AI-Eligibility Tagging:** Has every in-scope dataset been formally classified into Tier A (Open), Tier B (Private AI), Tier C (Masked Only), or Tier D (Prohibited)?	☐ Yes ☐ No
3	**Single Source of Truth:** Have duplicate spreadsheets and outdated document versions been archived so the system pulls exclusively from a verified Golden Record?	☐ Yes ☐ No
4	**Quality Baseline:** Has the target dataset been audited against the six quality dimensions (accuracy, completeness, consistency, timeliness, validity, uniqueness) within the last 90 days?	☐ Yes ☐ No
5	**Source Lineage:** Can a human operator trace any AI-generated output back to its exact underlying source record in under 60 seconds?	☐ Yes ☐ No
6	**Permission Inheritance:** Are technical controls in place ensuring the AI system cannot retrieve or display data that the prompting employee is not authorized to view?	☐ Yes ☐ No
7	**Automated PII Redaction:** Are DLP and masking guardrails active to strip regulated PII and sensitive commercial terms before data crosses an API boundary?	☐ Yes ☐ No
8	**Vendor Data Boundaries:** Do all third-party tool contracts include Zero Data Retention (ZDR), no-training clauses, and 30-day post-termination deletion guarantees?	☐ Yes ☐ No

#	Data Readiness Diagnostic	Response
9	**Named Human Stewardship:** Are a specific Executive Data Owner, Domain Data Steward, and Technical Custodian assigned to this dataset?	☐ Yes ☐ No
10	**Automated Drift & Quality Alerts:** Are automated checks running continuously to halt ingestion if upstream data schemas break or defect rates spike?	☐ Yes ☐ No

Interpreting Your Score

- **0 to 3 Points (Contaminated Perimeter):** Halt AI deployment immediately. Connecting models to this environment will amplify hallucinations, compliance exposure, and wasted compute. Execute Days 1 to 30 of the 90-Day Data Hygiene Sprint first.
- **4 to 7 Points (Partial Control):** Suitable for isolated, low-risk Tier 1 pilots only. Close remaining gaps in RBAC inheritance, lineage, and vendor contracts before scaling into customer-facing or financial workflows.
- **8 to 10 Points (AI-Ready Perimeter):** Green-Light Condition #2 is satisfied. Proceed to Phase 2 of the *180-Day "Build + Run" Roadmap* with monthly KPI tracking through your AI Governance Council.

Ready to build your organization's Clean Data Perimeter?

Schedule an executive briefing or 90-Day Data Hygiene Sprint with Artifiscal at **www.artifiscal.com**.