6/4/2025

AI Data Governance Checklist

“An Implementation Guide”

Morgan signing House 2025

1 “*AI Data Governance check list – A support to AI Implemenattion check List and Guide”*

Table of Contents

1. Introduction

2. Section 1: Data Availability & Quality

3. Section 2: Data Classification & Security

4. Section 3: Ethical AI & Compliance

5. Section 4: Data Stewardship & Lineage

6. Section 5: Continuous Data Governance

7. Instructions for Questionnaire Deployment

8. References

AI Data Governance Checklist & Implementation Guide

Inspired by "AI Assessment & Implementation Guide" (MSH 2025) and enriched with insights from global AI governance sources including ISO/IEC 42001, EU AI Act alignment, CAISO model, and data governance frameworks for generative and agentic AI.

Introduction

In AI governance, poor data quality isn't just a technical issue, it's a business risk. Inaccurate, incomplete, or biased data can distort model predictions, erode trust, and introduce regulatory exposure. This document provides a structured AI data governance framework with a focus on ensuring high-quality data and includes an actionable checklist and deployment guide.

Section 1: Strategic Data Alignment

1.1 **Mission Alignment**

- Are AI data initiatives aligned with the organization's digital transformation goals?
- Have data dependences for strategic AI use cases been identified?
- Does real data exist?
- Does data collection mechanism exist in the organization?

1.2 **Business Impact Mapping**

- Have data-driven AI use cases been prioritized by ROI, compliance risk, or customer value?

Section 2: AI Data Leadership & Oversight

2.1 **Executive Sponsorship**

- Is there a CAIO/CDO/CAISO accountable for data and AI governance?

2.2 **Governance Forums**

- Is a cross-functional Data & AI Governance Council in place (including Legal, Risk, IT, Ethics) to support program?
- Are policies ratified at Board or Subcommittee level?

Section 3: Data Architecture & Infrastructure Readiness

3.1 **Data Ecosystem Audit**

- Are all AI-relevant datasets cataloged with metadata, sensitivity, and ownership tags?
- Are data platforms aligned with FAIR principles (Findable, Accessible, Interoperable, Reusable)?

3.2 **Lineage & Observability**

- Are lineage tracking tools implemented (e.g., Apache Atlas, Collibra, Talend)?
- Are end-to-end lineage maps used to validate model inputs?

Section 4: AI Data Quality Governance Framework

4.1 **Why Data Quality Matters**

- Impacts model accuracy, fairness, and explainability.
- Influences compliance with global standards and legislation.
- Supports auditability and operational scaling.
- Enables scalable, auditable AI systems

4.2 **Key Dimensions of AI Data Quality**

Dimension	Description	Example Checklist Item
Accuracy	Data reflects real-world conditions	☑ Are invalid or placeholder values filtered out?
Completeness	No missing fields or critical gaps	☑ Are datasets ≥ 95% complete before training?
Consistency	Standardized schemas and units	☑ Are all fields normalized across platforms?

Timeliness	Data freshness and refresh cadence	☑ Are update intervals aligned with model expectations?
Uniqueness	Removal of duplicates and artifacts	☑ Are deduplication protocols embedded in pipelines?
Relevance	Fit-for-purpose data curation	☑ Is outdated or non-essential data pruned?

4.3 **Data Profiling & Readiness Checks**

- ☑ Are data profiling tools (e.g., Great Expectations) used routinely?
- ☑ Are validation rules and thresholds defined for key datasets?
- ☑ Is a retry/remediation process in place when quality drops?

4.4 **Dark Data & Non-Structured Sources**

- ☑ Are unstructured files mapped, OCR-processed, and classified?
- ☑ Is metadata attached for discoverability and governance?

4.5 **Quality Monitoring & Metrics**

Metric	Purpose
% Completeness	Ensure no field dropout
Data Drift Score	Flag distributional shifts
Timeliness Index	Assess data freshness
Validation Pass Rate	Track rule-based data fitness

4.6 **Tools & Dashboards**

Tool	Function
Monte Carlo	Data observability
Talend	ETL pipeline & transformation
Collibra	Governance & stewardship
SHAP/LIME	Model transparency via data

Section 5: Data Security & Classification for AI

5.1 **Access Controls**

- Is RBAC or ABAC enforced on all data sources for AI use?
- Are regular access reviews and entitlement audits conducted?

5.2 **Sensitive Data Handling**

- Is PII/PHI masked, tokenized, or anonymized prior to model training?
- Is end-to-end encryption (TLS1.2+, AES-256) in place for data in transit and at rest?

Section 6: Ethical Data Use & Compliance Alignment

6.1 **Ethics Board Review**

- Are high-risk AI use cases data subject to formal Ethics Board review?
- Are bias data audits conducted pre- and post-deployment?

6.2 **Compliance Mapping**

- Is each AI model mapped to applicable standards (e.g., ISO/IEC 42001, ISO/IEC 27001, GDPR, CCPA, EU AI Act)?
- Are automated checks in place for regulatory policy triggers to check data?

Section 7: Stewardship, Ownership & Lineage Governance

7.1 **Defined Roles**

- Are data owners and stewards assigned to each dataset and model?
- Are responsibilities tracked in a RACI matrix?

7.2 **Auditability**

- Are immutable logs retained for data access, transformation, and model output events?

Section 8: Data Risk Management

8.1 **Risk Register**

- Is there a centralized register tracking data quality, privacy, lineage, and security risks?

- Are data-related risks scored, and mitigation plans documented?

8.2 **Incident Readiness**

- Are data-related AI breaches covered in the organization's IR playbook?

Section 9: Model Transparency & Data Traceability

9.1 **Explainability Assets**

- Are Data Sheets and Model Cards available for every production model?

9.2 **Traceback Protocols**

- Can each model output be traced back to the data source, transformation, and version?

Section 10: Continuous Data Governance Improvement

10.1 **Lifecycle Monitoring**

- Are metrics and logs continuously monitored for drift, degradation, and risk?
- Are retraining triggers based on data quality thresholds?

10.2 **Policy Refresh Cadence**

- Are data governance policies reviewed at least every 6 months?
- Are updates communicated via dashboards and stakeholder briefings?

Appendix: Maturity Markers for AI Data Governance

- Level 1: Ad Hoc – No clear ownership or governance.
- Level 2: Defined – Basic roles and processes exist.
- Level 3: Operationalized – Cross-functional governance in action.
- Level 4: Measured – KPIs tracked, audits performed.
- Level 5: Adaptive – Continuous improvement, policy agility, regulatory alignment.

Instructions for Questionnaire Deployment

- **Question Types:** For each question, determine whether it is:
  - Yes/No (e.g., “Is there an AI Ethics Board?”),
  - Multiple Choice (e.g., “Which AI frameworks are considered? [TensorFlow, PyTorch, Scikit-Learn, Other]),
  - Rating Scale (e.g., “Rate our data quality: 1 = Poor, 5 = Excellent”), or
  - Open-Ended (e.g., “Describe the primary AI objectives for your business unit”).

• **Question Types:** Use Yes/No, Multiple Choice, Rating Scale, or Open-Ended formats.

• **Survey Logic:** Use skip logic (e.g., if 'No' to CAISO, skip to alternate responsibility).

• **Required vs. Optional:** Mark sections like Data Security and Risk Governance as 'Required'.

• **Sections & Progress Bar:** Organize by topic and display progress to respondents.

• Distribution: Send to all relevant stakeholders (e.g., IT, compliance, exec sponsors).

• Living Document: Revise based on responses and feedback to fill governance gaps.

Use this questionnaire as a living document: iterate after initial responses, refine questions to address gaps, and ensure every area—Strategy, Leadership & Governance, Technology, Data Governance, Security & Risk, Talent & Change, Pilot & MLOps, KPI Monitoring, and Board Oversight—is thoroughly covered before embarking on or scaling AI initiatives.

References

- Morgan Signing House (2025). AI Assessment & Implementation Guide.
- Pullum, S. (2025). CONVERGENCE. The CAISO. Copenhagen Compliance.

- Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction Machines.

- Advisera. How to Handle Artificial Intelligence Threats Using ISO 27001.

- Patrick Sullivan. Bridging the Gaps – ISO/IEC 42001 and EU AI QMS.

- Risk3Sixty. Step-by-Step Guide to ISO 42001 Certification.

- SSRN. Risks and Limitations of Generative AI Systems. SSRN ID: 5191295