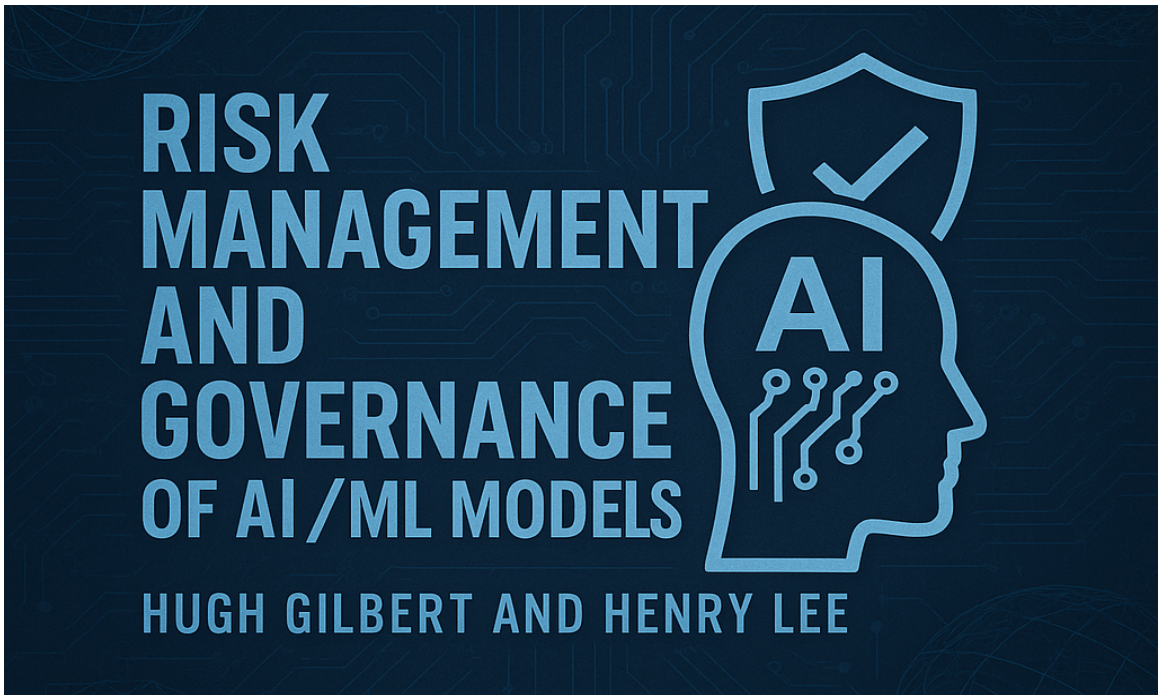

Risk Management and Governance of AI / ML Models

In order to promote thought leadership in the discipline, we offer the following series of articles. The topic of this particular post is highlighted, followed by the article.

Hugh Gilbert and Henry Lee

4.4 Data Governance as a Pillar of AI Risk Management

Introduction

For decades, it has been a truism among developers and users of models that a model is only as good as the data that serves as its raw material: GIGO (Garbage-in, Garbage-out) goes the inveterate adage. And *a fortiori*, the comparatively opaque Artificial intelligence (AI) and Machine Learning (ML) models of today are only as trustworthy as the data on which they are built and operated. In financial institutions, data feeds and pipelines not only drive model predictions but also undergird regulatory reporting, customer treatment, fraud controls, and capital decisions. Weaknesses in data quality, lineage, or integrity translate directly into model risk. As AI/ML usage grows across business lines, data governance is no longer merely a parallel discipline, an adjunct to the nontrivial work of conceptual soundness and developmental evidence, but rather a core pillar of Model Risk Management (MRM) and broader AI governance.

For risk management professionals, this means casting a wide frame beyond algorithms and validation techniques to the full data lifecycle: sourcing, ingestion, transformation, storage, usage, and retirement. This document explores how data governance supports AI/ML model risk management, the specific challenges institutions face, and practical approaches to mitigating data-driven risks.

1. Why Data Governance is Central to AI/ML Model Risk

Traditional MRM frameworks focus on conceptual soundness, process controls, and model performance monitoring. In AI/ML, however, the domains of data and model cannot be regarded as disjoint, non-overlapping sets. Complex models can memorize problematic patterns in data, learn spurious correlations, or embed historical biases such that data issues manifest as model vulnerabilities: bias, instability, overfitting, and lack of robustness to changing conditions.

Sound data governance provides the foundation for trustworthy models by ensuring that data is accurate, complete, relevant, timely, and used in a manner consistent with legal, ethical, and policy expectations. Without this foundation, even the most technically sophisticated models cannot be considered as well-governed.

2. Key Data Risks in AI/ML Model Risk Management

Risk professionals should be able to articulate and assess specific data risks that affect AI/ML models. Key categories include the following.

2.1 Data Quality and Integrity

Poor data quality, such as missing values, inconsistent identifiers, incorrect labels, and duplicate records, directly degrades model performance and may render validation results unreliable. That is to say, the validation is conditioned on poor-quality data and is likely not to generalize well with new data. In supervised learning, label errors can be particularly damaging, as models learn from incorrect outcomes. Integrity issues, such as untracked manual overrides or undocumented transformations, make it difficult to reproduce results or trace anomalies.

2.2 Representativeness and Bias

Data that does not adequately represent the population to which a model will be applied creates significant risk. Under-representation of specific customer segments, geographies, or economic regimes can lead to biased decisions and fairness concerns, especially in credit, pricing, and fraud models. Historic data may embed discriminatory patterns that AI models amplify unless explicitly mitigated.

2.3 Lineage and Provenance

Without clear data lineage, the knowledge of where data originated, how it has been transformed, and which models depend on it, institutions struggle to assess the impact of upstream changes or errors. Lack of provenance also undermines regulatory attestations, such as explaining how inputs to capital or liquidity models are derived.

2.4 Privacy, Confidentiality, and Consent

AI/ML models frequently use granular customer data, transaction history, behavioral signals, and external data sources. Misalignment with privacy regulations, consent obligations, or internal data protection policies creates legal and conduct risk. Improperly governed data sharing with external vendors or cloud platforms further heightens the stakes.

2.5 Access Control and Segregation of Duties

If data used for models can be accessed or modified by unauthorized parties, the institution faces both security and model integrity risks. Inadequate segregation of duties, such as the same person controlling both data preparation and model approval, weakens assurance and may conflict with regulatory expectations for independent review.

2.6 Timeliness and Volatility

For many AI/ML models, particularly those used in trading, fraud, and real-time decisioning, stale data can be as dangerous as inaccurate data. Misaligned refresh frequencies between data sources and models may cause predictions to lag current

conditions. Conversely, highly volatile data streams without appropriate smoothing or controls can create instability and false positives.

2.7 Third-Party and Alternative Data

Growing use of third-party data and alternative data sources introduces risks related to licensing, provenance, ongoing availability, and quality. Risk professionals must understand not only how such data is used in models but also the contractual and ethical constraints governing it.

3. Data Governance Framework for AI/ML Model Risk

A robust data governance framework aligns people, processes, and technology to control data risk across the model lifecycle. For AI/ML, the framework should be tightly integrated with MRM and operational risk management processes.

3.1 Roles and Accountability

Key roles typically include:

- **Chief Data Officer (CDO) and Data Governance Office** – Define enterprise data policies, standards, and governance forums. Coordinate across business units.
- **Data Owners and Stewards** – Own specific data domains and are accountable for quality, lineage, and access controls.
- **Model Owners** – Responsible for appropriate use of data within models, conformance with policy, and documentation of assumptions and limitations.
- **Model Risk Management and Independent Validators** – Challenge whether data used is fit for purpose, appropriately governed, and adequately documented.
- **Technology and Security Teams** – Implement platforms and controls for storage, access, encryption, and monitoring.

Clear RACI (Responsible, Accountable, Consulted, Informed) structures ensure that data issues have identifiable owners and that model risk stakeholders can ascribe responsibility to the right parties.

3.2 Policies, Standards, and Control Libraries

Policies translate high-level risk appetite into actionable rules for data usage. For AI/ML models, institutions typically define standards covering:

- Data quality thresholds and remediation expectations
- Approved sources for critical data elements
- Retention periods and archival requirements
- Requirements for capturing lineage and metadata

- Privacy, consent, and anonymization practices

These standards should be supported by a library of data controls, both detective and preventive, that can be applied consistently across projects. Examples include input validation rules, reconciliation checks, segregation of environments, encryption requirements, and approval workflows for introducing new data sources.

3.3 Metadata Management and Data Catalogs

At AI scale, manual tracking of data assets is not viable. Data catalogs provide a central view of datasets, their business definitions, owners, quality metrics, and usage in downstream models. Integrated with model inventories, catalogs help risk teams quickly answer questions such as:

- Which models rely on this data source?
- Which products, customers, or reports are affected if quality issues are discovered?
- Are there conflicting versions of the same data element being used in different models?

A well-maintained catalog becomes an enabler for both governance and reuse, reducing duplication of effort and inconsistencies.

3.4 Data Lineage and Impact Analysis

Technical lineage tools capture how data flows through ingestion, transformation, feature engineering, and model scoring pipelines. When combined with business process maps, this lineage enables impact analysis for upstream changes such as a system decommission, a new vendor relationship, or redefinition of a data element.

For model risk management, such capabilities are invaluable in understanding whether a model breach is due to conceptual weaknesses or data pipeline issues. If stakeholders ask “where did this number come from?”, lineage diagrams allow institutions to respond quickly and informatively.

3.5 Data Quality Monitoring and Issue Management

Effective data governance requires ongoing measurement. Automated rules and statistical checks can monitor completeness, validity, consistency, and distributional stability of key variables. Alerts feed into issue management workflows where data stewards and model owners jointly:

- Assess severity for triage and potential business impact
- Define remediation actions and timelines
- Decide whether compensating controls or temporary thresholds are required
- Document residual risk and communicate to relevant governance forums

In those instances in which data quality cannot be fully remediated, model documentation should clearly state limitations and any adjustments (such as conservative overlays or tighter usage thresholds) applied to compensate for these weaknesses.

4. Embedding Data Governance Across the AI/ML Model Lifecycle

Data governance is most effective when embedded into each stage of the model lifecycle rather than treated as a one-time gate. Risk professionals should ensure that the following practices are in place.

4.1 Problem Framing and Data Selection

During model ideation, teams should explicitly consider which data is necessary and whether its use is lawful, ethical, and aligned with customer expectations. Early involvement of data owners, privacy officers, and compliance helps avoid downstream rework and reduces the likelihood of models being built on prohibited or sensitive attributes, such as protected characteristics or opaque alternative data.

4.2 Development and Feature Engineering

In development, data governance focuses on reproducibility and suitability. Feature engineering pipelines should be codified rather than created through one-off manual steps. Sampling methods, outlier treatment, and exclusions must be fully documented. Fairness testing should be undertaken to ensure that customer outcomes are consistent with guidelines and legal requirements, and results are recorded as part of the model development documentation.

Synthetic or augmented data (e.g., interpolation) should be clearly labeled and assessed for realism and bias. If synthetic data is used to overcome sparsity in certain subpopulations, governance must evaluate whether it genuinely improves fairness or introduces artefacts that inequitably distort model behavior.

4.3 Validation and Independent Review

Validators should challenge both the **choice** and **treatment** of data. Typical questions may include:

- Is the observation window consistent with the business process being modeled?
- Are there data leakage risks whereby future information leaks into training features?
- Are model results unduly sensitive to specific data sources, time periods, or preprocessing steps?
- Do alternative samples (e.g., out-of-time validation, stressed subsamples) materially change performance?

Independent numerical verification of key performance metrics using raw data is a powerful control, as it tests both the robustness of the modelling process and the reliability of the underlying data.

4.4 Deployment, Monitoring, and Change Management

Post-deployment, monitoring regimes should track not only model outputs but also input data quality and coverage. Examples include:

- Thresholds for missing data or unexpected category levels
- Drift metrics for key features relative to training distributions
- Monitoring of upstream system changes and vendor data updates

Material changes to data sources, transformation logic, or external vendors should follow formal change management processes with impact analysis and, where necessary, re-validation of affected models. Data-related incidents, such as corrupted feeds or mis-mapped fields, should be logged, investigated, and linked back to model risk registers and operational incident reporting.

4.5 Retirement and Data Archival

When models are decommissioned or replaced, data governance considerations remain. Institutions should determine what training and scoring data must be retained for audit purposes, how long it should be kept, and how it should be protected or anonymized.

Clear archival practices help respond to retrospective regulatory queries about past decisions driven by AI models (for example, explaining why a loan was declined two years earlier) and support forensic analysis if historic issues are discovered.

5. Practical Challenges in Implementing Data Governance for AI/ML

Even where the importance of data governance is widely recognized, institutions face substantial practical obstacles in operationalizing it for AI/ML:

- **Legacy and siloed data architectures.** Many banks operate with fragmented data landscapes (e.g., conversion of legacy systems), making end-to-end lineage and quality monitoring difficult to achieve.
- **Speed versus control.** Business lines may regard governance processes as inhibitory to timely delivery of AI use cases, leading to pressure for exceptions or parallel “shadow” data pipelines.
- **Limited ability to quantify data risk.** Unlike credit or market risk, data risk metrics may be less standardized, which complicates incorporation into risk appetite and capital frameworks.

- **Third-party and cloud dependencies.** Outsourced data feeds and cloud platforms complicate visibility into data handling practices, resilience capabilities, and incident response.
- **Cross-jurisdictional regulations.** Global institutions must reconcile overlapping and sometimes conflicting privacy, localization, and secrecy requirements across regions.

Recognizing these challenges openly helps risk management professionals design pragmatic, phased approaches instead of aiming for perfection from the outset.

6. Mitigation Strategies and Good Practices

To address these challenges, risk management professionals can champion a pragmatic set of mitigation strategies that align data governance with AI/ML objectives. Key levers include:

- **Define a data risk appetite.** Articulate qualitative and quantitative statements about acceptable levels of data quality issues, lineage gaps, and unresolved incidents for different model tiers. This frames prioritization and investment decisions.
- **Integrate data controls into model policies.** Ensure model development and validation standards explicitly reference required data governance artefacts, such as data dictionaries, lineage diagrams, quality reports, and fairness analyses.
- **Invest in enabling tooling.** Prioritize data catalogs, lineage tools, and automated quality monitoring capabilities that provide reusable utilities across multiple models, rather than building bespoke solutions per use case.
- **Standardize "data packs" for validators.** Require model owners to provide structured data documentation packages, including source descriptions, sampling logic, known limitations, and fairness analysis, to streamline independent review and reduce back-and-forth queries.
- **Promote training and culture.** Educate data scientists, product owners, and risk teams on how data decisions translate into model risk. Case studies of real-world failures are particularly effective in building intuition and buy-in.
- **Use metrics and key risk indicators (KRIs).** Track indicators such as the percentage of critical data elements with assigned owners, the number of unresolved data quality issues affecting high-risk models, the share of models with documented lineage, and the frequency of data-related incidents. Over time, these metrics can be linked to operational loss events and remediation outcomes.

Conclusion

Data governance is not an auxiliary function but a **foundational pillar** of AI risk management. In the context of financial institutions, wherein AI/ML models increasingly drive decisions with regulatory, financial, and ethical implications, weak control over data equates to potential lapses in model governance.

By clarifying accountabilities, strengthening policies and tooling, and embedding data considerations into every stage of the model lifecycle, institutions can materially reduce the probability and impact of AI-related failures. For risk management professionals, engaging deeply with data governance is an opportunity to move from reactive model remediation to dynamic risk prevention.

Institutions that treat high-quality, well-governed data as a strategic asset will be better positioned to deploy AI safely, producing quality output while maintaining trust with customers, regulators, and other stakeholders.