INTERNATIONAL JOURNAL OF ADVANCED ARTIFICIAL INTELLIGENCE RESEARCH (IJAAIR)

eISSN: 3087-4297

Volume. 03, Issue. 04, pp. 26-32, April 2026"

Data Architecture Maturity as A Predictor of Enterprise AI Success in Regulated Industries

Mohammed Arbaaz Shareef
Lead Data Engineer at Anblicks

Article Received: 11/01/2026, Article Revised: 07/02/2026, Article Accepted: 20/03/2026, Article Published: 18/04/2026
DOI: https://doi.org/10.55640/ijaair-v03i04-03

© 2026 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the Creative Commons Attribution License 4.0 (CC-BY), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

ABSTRACT

Objective: To operationalize data-architecture maturity as a predictor of enterprise AI success in regulated industries, where auditability, traceability, and lifecycle risk controls constrain deployment at scale.

Methods: A structured comparative synthesis of ten recent peer-reviewed and standards sources on data governance and quality, maturity modeling, lakehouse/platform capabilities, MLOps, and AI risk management was performed, followed by construct-to-indicator mapping that links maturity domains to observable operational signals.

Results: The synthesis yields a five-domain predictor framework—(1) governance maturity, (2) data quality maturity, (3) platform architecture maturity, (4) MLOps maturity, and (5) responsible-AI maturity—paired with an evidence-linked mapping to success mechanisms (reproducibility, controllability, deployment stability, auditable decision records, and sustained performance under drift). The framework is made assessable pre-deployment through a concise signal set: versioned datasets and reconstructable training snapshots, transformation lineage bound to metadata, continuous quality metrics with documented thresholds, automated drift monitoring with traceable alerts, and explicit lifecycle risk controls aligned with regulatory expectations.

Implications: The model supports feasibility screening and targeted remediation planning before scaling AI in compliance-constrained environments.

KEYWORDS

data architecture maturity, enterprise AI, regulated industries, data governance, data quality, metadata and lineage.

Introduction

Regulated industries face a structural constraint: AI value depends not only on model accuracy, but also on the organization's ability to demonstrate the origin of data, its justification for use, the process by which it was transformed, and the mechanisms for monitoring model behavior over time. In such environments, weaknesses in architecture—such as fragmented data domains, inconsistent semantics, missing lineage, and non-repeatable pipelines—convert AI delivery into a sequence of exception-handling episodes rather than an engineering process. The resulting gap between "AI prototypes" and "enterprise AI" motivate the present study.

The purpose of the study is to conceptually substantiate data architecture maturity as a structural determinant of enterprise AI success in regulated industries, where compliance, auditability, and lifecycle control constrain AI deployment.

To achieve this purpose, the study addresses the following objectives:

1) to systematize data architecture maturity domains relevant for AI readiness in regulated organizational contexts;

2) to establish explanatory links between these maturity domains and observable mechanisms of AI success, including reproducibility, operational stability, governance alignment, and sustained performance;

3) to develop an integrated maturity-based predictor framework that enables pre-deployment assessment of enterprise AI feasibility.

Scientific novelty is associated with a unified maturity-to-success mapping that connects data-management maturity structures with AI governance and operationalization practices, allowing AI outcomes to be interpreted as an architectural consequence rather than a purely modeling consequence.

Materials and Methods

The evidence base was formed from ten recent sources selected for direct relevance to maturity models, data governance and quality, AI operationalization, and modern enterprise data platforms: T. Alexakis [1] synthesized standards and methodologies for assessing and managing data quality; B. M. V. Bernardo [2] reviewed data governance and data-quality management, including governance frameworks and maturity approaches; P. Choowan [3] analyzed AI-enabled financial decision-making and quantified governance maturity as a mediator in AI-to-outcome relations; B. Eken [4] synthesized MLOps practices and adoption challenges across peer-reviewed and gray literature; R. M. F. D. Ferreira [5] piloted a maturity model for responsible AI in organizational settings; A. A. Harby [6] compared lake, warehouse, and lakehouse systems and proposed an advanced lakehouse architecture; D. Mazumdar [7] decomposed lakehouse architecture into implementation-independent components and capabilities; NIST [8] formalized a risk-management framework for AI systems across lifecycle functions; A. Paleyes [9] surveyed deployment challenges across the ML workflow; L. Pörtner [10] developed and validated a data management maturity model with process dimensions and capabilities relevant for data-driven organizations.

The study employed non-experimental analytic methods, including a comparative analysis of conceptual models, a structured synthesis of findings across sources, and the mapping of constructs into an integrated explanatory framework. The approach combined concept-to-indicator mapping (maturity domains → operational signals), cross-source triangulation (governance, quality, platform, MLOps, and risk), and interpretive analysis focused on the constraints of the regulated industry.

The selection of sources was guided by three criteria:

1) explicit engagement with maturity models, governance, or lifecycle control of data and AI systems;

2) relevance to enterprise-scale or regulated organizational contexts;

3) publication in peer-reviewed journals or authoritative institutional frameworks.

Although the evidence base is intentionally limited in size, it reflects conceptual diversity rather than statistical representation, allowing for cross-domain triangulation between governance, platform architecture, MLOps, and responsible AI research.

Results

Across the reviewed literature, maturity constructs converge on a recurring idea: enterprise AI performance and sustainability depend on whether data is governed, intelligible, controllable, and reproducible within an operational platform that supports both analytics and lifecycle controls. Data architecture maturity, in this framing, is not reducible to a single “platform choice”; it is a composite capability spanning governance structures, quality discipline, metadata/lineage, architectural integration, and operationalization routines that keep models aligned with data drift, policy drift, and organizational accountability needs. The maturity-model literature indicates that organizations typically represent progress in staged levels and assess readiness through capability sets rather than isolated technologies. In AI maturity model research, maturity grids and multi-level representations remain prevalent, and models often separate domains such as strategy, data, technology, people, and governance to explain why AI adoption outcomes differ between organizations [1].

A central empirical signal supporting the predictive function of governance maturity emerges from financial

decision-making research. In the systematic review by Choowan and Daovisan, governance maturity was quantified as a mediating construct connecting AI integration to financial decision outcomes; the reported structural relationships indicate that algorithmic integration alone does not translate into decision-quality gains when governance mechanisms are insufficiently mature, while stronger governance links AI capability to measurable outcomes [3]. This aligns with broader governance syntheses that treat governance not as documentation overhead but as an organizing system for data ownership, quality accountability, and control of permissible use, with maturity models used to diagnose gaps and structure improvement pathways [2]. For regulated industries, the predictive value of data architecture maturity becomes apparent precisely at this junction: when governance maturity is low, AI systems inherit ambiguity in provenance, entitlement, and semantic consistency, thereby increasing the cost of validation and restricting the deployment scope. When governance maturity is higher, AI delivery becomes more repeatable because control objectives are embedded into data flows and operating routines rather than enforced through ad hoc reviews.

Data quality discipline contributes a second predictor channel. Data quality research emphasizes that quality is operationally multidimensional and tied to formal standards and measurement methodologies. The applied implication for enterprise AI is that quality cannot remain a pre-training checklist, because production AI depends on continuous quality assurance, monitoring, and revalidation of data products used by models [1]. This is suitable for regulated settings where traceability and demonstrable control over data transformations are crucial in determining whether a model can be approved for use in core processes. In an architecture maturity interpretation, quality maturity is evident through standardized metrics, systematic profiling, validation routines, and tooling support that can be audited. Weak quality maturity amplifies hidden failure modes: models are trained on datasets that cannot be reliably reconstructed, feature definitions drift across teams, and post-deployment incidents become difficult to attribute to either data changes or model changes.

A third predictor channel is platform architecture maturity, increasingly discussed through the lakehouse paradigm. Lakehouse research decomposes the architecture into capabilities such as transactional consistency, open data formats, reduced copy proliferation, and governance layers on top of data-lake storage [7]. Complementary survey evidence compares lakehouse systems with lakes and warehouses, highlighting that the lakehouse aims to unify storage flexibility with warehouse-like management and performance characteristics [6]. From a maturity perspective, this matters because regulated enterprise AI repeatedly needs repeatable training data snapshots, consistent feature extraction, and controlled write paths across teams. Architectural characteristics, such as transactional support and table-format governance, are relevant insofar as they enable consistent reads and writes, as well as stable dataset versioning, which support reproducibility and audit trails in model development and monitoring [7]. Case-study evidence from industrial lakehouse implementations reinforces that architectural decisions and operational experiences become decisive for AI applications at scale, with data platforms treated as core enablers for analytics and AI workloads [10]. In regulated industries, the same logic applies: platform maturity predicts whether AI teams can produce defensible evidence chains for training datasets and transformations, as well as whether lineage is accessible for audit and incident analysis.

Operationalization maturity forms a fourth predictor channel and is closely tied to MLOps. A multivocal synthesis of MLOps reveals recurring practices and challenges throughout the ML lifecycle, encompassing pipeline complexity, artifact and environment management, monitoring, governance, and socio-technical coordination across teams [4]. In parallel, the survey of deployment case studies highlights that challenges arise at every stage of deployment workflows, suggesting that success depends on how models are integrated into data ingestion, validation, monitoring, and downstream systems, rather than solely on modeling [9]. The combined interpretation is that architecture maturity predicts AI success through operational stability: mature organizations institutionalize versioning of data and artifacts, monitoring for drift, controlled releases, and traceable decision records, while less mature organizations rely on manual handoffs and brittle dependencies that break under regulatory scrutiny and change.

Responsible AI maturity introduces a fifth predictor channel that connects architecture maturity with governance credibility. Evidence from the piloting of a responsible AI maturity model suggests that organizations may recognize responsible AI requirements yet respond reactively, and maturity assessments help reveal gaps and needs for more

trustworthy development and deployment practices [5]. In regulated industries, such maturity is not separable from data architecture: fairness, privacy protection, explainability documentation, and accountability routines depend on whether data lineage, access governance, and reproducible pipelines exist in practice. The NIST AI Risk Management Framework formalizes lifecycle-oriented functions (governance, contextual mapping, measurement, and management) that emphasize organizational processes for managing AI risks and maintaining trustworthiness [8]. When mapped to data architecture maturity, these functions translate into concrete architectural demands: controlled data use, recorded transformations, measurable performance and risk indicators, and operational controls that continue after deployment.

On this basis, data architecture maturity can be expressed as an integrated predictor model with five domains that jointly explain enterprise AI success under regulation: governance maturity (ownership, policy enforcement, audit trails), quality maturity (metrics, validation routines, standard alignment), platform maturity (transactional consistency, open formats, reduced duplication, governed tables), MLOps maturity (artifact management, monitoring, controlled releases, cross-team workflows), and responsible AI maturity (risk mapping, measurement, lifecycle governance). Distinct evidence streams support each domain, yet they reinforce one another in practice: governance without platform support remains bureaucratic; platform capability without MLOps does not create stable operations; MLOps without quality discipline automates unreliable inputs; responsible AI objectives lack enforcement without governance and lineage.

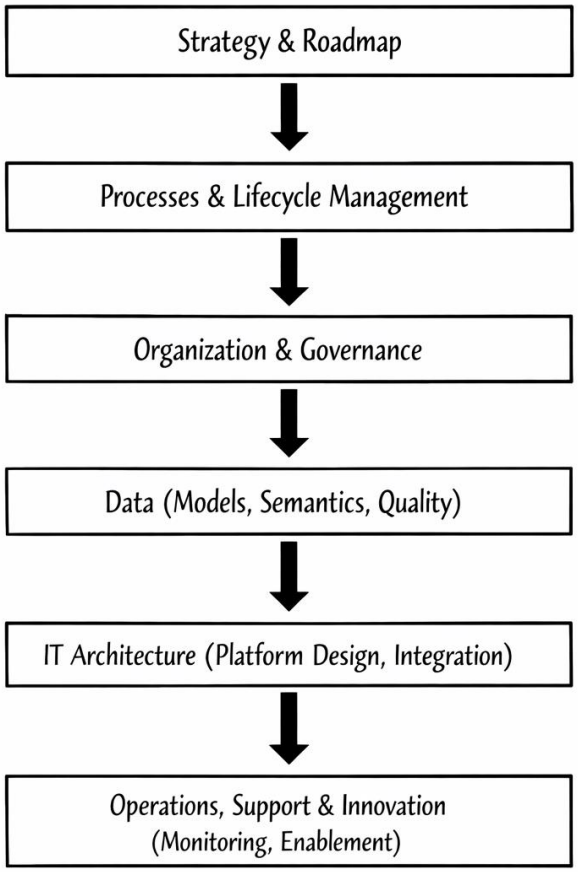


Figure 1. Data-management maturity pillars relevant for AI readiness (adapted from Pörtner et al. [10])

The pillars operationalize what "maturity" means in a way that can be assessed before AI scaling decisions. When these pillars are interpreted through the lens of regulated industry demands, AI success becomes predictable: higher maturity implies stronger repeatability, more transparent accountability, reduced ambiguity in data interpretation, and more reliable operating routines, which collectively reduce the friction of compliance, validation, and incident response.

Discussion

The synthesized results support a practical interpretation: in regulated industries, "AI success" is structurally constrained by the organization's ability to maintain evidence chains across data, models, and decisions. The governance literature positions maturity

models as tools to formalize practices, diagnose gaps, and guide improvement trajectories rather than as abstract scoring exercises [2]. The financial evidence strengthens this position by demonstrating that governance maturity mediates the value pathway between AI integration and decision outcomes, implying that architecture maturity has predictive power precisely because it captures the organization's capacity to convert algorithmic capability into controlled and accountable decision support [3]. The risk-management viewpoint further clarifies why regulated industries cannot treat AI as a model-only problem: lifecycle governance and risk measurement functions require stable data foundations and operational controls that persist after deployment [8].

Table 1. Evidence-linked mapping between maturity domains and AI success mechanisms in regulated industries [1–10]

Maturity domain	What becomes predictable at higher maturity	Evidence basis
Governance maturity	Lower compliance friction through clearer ownership, entitlement, and auditable control routines	Governance frameworks and maturity models; governance maturity mediating outcomes in finance
Data quality maturity	More stable model behavior and fewer unexplained performance regressions due to measurable, standardized quality management	Data quality standards and methodologies; governance-quality coupling
Platform architecture maturity	Higher reproducibility via governed tables, consistent read/write semantics, and reduced copy proliferation	Lakehouse capability decomposition; lakehouse survey and comparative analysis; real-world lakehouse architecture experiences
MLOps maturity	Shorter path from development to controlled deployment and improved monitoring/rollback discipline	MLOps synthesis of practices and challenges; deployment workflow challenges across stages
Responsible AI maturity	More defensible approvals and sustained trust due to institutionalized requirement tracking and gap discovery	Responsible AI maturity piloting; lifecycle risk functions in AI RMF

A mature architecture narrows uncertainty: the organization can reconstruct training data, justify feature definitions, document decision logic, and monitor drift with traceable evidence. A less mature architecture expands uncertainty: even accurate models become operationally fragile, harder to validate, and riskier to deploy in core regulated processes.

Deployment and lifecycle findings further sharpen the discussion. The deployment survey shows that barriers accumulate across the workflow—from data ingestion and pipeline integration to monitoring and cross-cutting concerns such as security and trust—so success depends on systemic engineering rather than local optimization [9]. The MLOps synthesis notes that adoption challenges are socio-technical, as organizations struggle with pipeline complexity, artifact management, and governance integration. Therefore, maturity must be interpreted as an organizational capability to standardize workflows and enforce controls through tooling and collaboration routines [4]. In regulated industries, these constraints become more binding because audit and compliance requirements often demand explicit evidence for data provenance, model validation, and operational monitoring.

Table 2. Operational signals used to assess architecture maturity as an AI-success predictor [1; 2; 4; 5; 7; 8–10]

Operational signal observable in practice	Why does titpredicts AI success under regulation	Evidence basis
Versioned datasets and reconstructable training snapshots	Supports reproducibility, auditability, and defensible validation of model changes	Lakehouse transactional/governance capabilities; MLOps artifact and pipeline control focus
Documented lineage and metadata tied to transformations	Reduces ambiguity in provenance and accelerates incident/root-cause analysis	Governance and maturity model emphasizes governance/metadata; maturity pillars covering data and IT architecture
Automated monitoring for data and model drift with traceable alerts	Converts post-deployment risk into measurable signals and controlled responses	MLOps monitoring and lifecycle practices; deployment-stage challenges across workflow
Standardized quality metrics are applied continuously	Prevents silent degradation and supports consistent compliance reporting	Data-quality standards and methodologies; governance-quality linkage
Explicit lifecycle risk governance mapped to organizational controls	Aligns AI operation with risk functions expected in regulated settings	AI RMF lifecycle governance and measurement functions; responsible for AI maturity gap identification

The platform literature strengthens this argument by showing that architectural capabilities (transactional support, open formats, governed tables) are not abstract virtues; they support practical outcomes such as safe concurrent access, reduced duplication, and more reliable governance, which directly influence whether AI pipelines can be repeated and validated over time [7].

A limitation of the present analytic design is the absence of a new dataset linking maturity scores to longitudinal AI outcomes within a single regulated sector. Yet the selected evidence base supports defensible, non-speculative inferences: governance maturity empirically mediates AI value in finance [3]; deployment and MLOps syntheses show recurring failure modes that maturity practices are designed to mitigate [4; 9]; quality and governance literatures provide operational methods for measurement rather than rhetorical claims [1; 2]. A next step for empirical extension would be a multi-organization study that operationalizes the signals in Table 2 into measurable indicators and tests their predictive association with deployment stability, audit outcomes, and post-deployment incident rates.

Conclusion

The study systematized maturity domains that describe data architecture readiness for enterprise AI in regulated industries and linked them to mechanisms that determine AI success under compliance constraints. The synthesis supports the first task by consolidating maturity into five domains—governance, quality, platform architecture, MLOps, and responsible AI—grounded in recent research on maturity modeling, data governance, lakehouse architectures, ML deployment, and AI risk management. The second task was met by showing how each domain predicts distinct components of success: governance maturity conditions whether AI capability translates into decision outcomes in finance; quality maturity stabilizes model behavior through measurable standards; platform maturity enables reproducibility through governed data structures; MLOps maturity sustains lifecycle control; responsible AI maturity translates requirements into actionable gap discovery and controls. The third task was addressed through an integrated predictor model operationalized via observable signals (dataset versioning, lineage, monitoring, standardized quality metrics, lifecycle risk governance), offering a basis for pre-rollout feasibility assessment and targeted maturity improvement before scaling AI in regulated environments.

By framing AI success as an architectural consequence rather than a purely modeling outcome, the study contributes a diagnostic perspective that supports more informed investment, sequencing, and risk management

decisions in enterprise AI programs operating under regulatory constraints.

References

1. Alexakis, T., Adamopoulou, E., Peppes, N., Daskalakis, E., & Ntouskas, G. (2025). Evaluating data quality: Comparative insights on standards, methodologies, and modern software tools. Electronics, 14(15), 3038. https://doi.org/10.3390/electronics14153038

2. Bernardo, B. M. V., Mamede, H. S., Barroso, J. M. P., & Santos, V. M. P. D. D. (2024). Data governance & quality management—Innovation and breakthroughs across different fields. Journal of Innovation & Knowledge, 9(4), 100598. https://doi.org/10.1016/j.jik.2024.100598

3. Choowan, P., & Daovisan, H. (2026). Artificial intelligence in data governance for financial decision-making: A systematic review. Big Data and Cognitive Computing, 10(1), 8. https://doi.org/10.3390/bdcc10010008

4. Eken, B., Pallewatta, S., Tran, N., Tosun, A., & Babar, M. A. (2025). A multivocal review of MLOps practices, challenges and open issues. ACM Computing Surveys, 58(2), 1–35.

5. Ferreira, R. M. F. D., Grilo, A., & Maia, M. (2025). Piloting a maturity model for responsible artificial intelligence: A Portuguese case study. Journal of Responsible Technology, 22, 100117. https://doi.org/10.1016/j.jrt.2025.100117

6. Harby, A. A., & Zulkernine, F. (2025). Data lakehouse: A survey and experimental study. Information Systems, 127, 102460. https://doi.org/10.1016/j.is.2024.102460

7. Mazumdar, D., Hughes, J., & Onofre, J. B. (2023). The data lakehouse: Data warehousing and more. arXiv. https://arxiv.org/abs/2310.08697

8. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

9. Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: A survey of case studies. ACM Computing Surveys, 55(6), Article 1, 29. https://doi.org/10.1145/3533378

10. Pörtner, L., Riel, A., Schmidt, B., Leclaire, M., & Möske, R. (2025). Data management maturity model—Process dimensions and capabilities to leverage data-driven organizations towards Industry 5.0. Applied System Innovation, 8(2), 41. https://doi.org/10.3390/asi8020041