E-ISSN: 2655-0865
Ranah Research
Journal of Multidisciplinary Research and Development
082170743613 ranahresearch@gmail.com https://jurnal.ranahresearch.com
DINASTI RESEARCH

DOI: https://doi.org/10.38035/rrj.v8i4

Data Governance Practices for Risk Management in Artificial Intelligence Implementation: A Systematic Literature Review

Adib Prima Yadri[1], Rizal Fathoni Aji[2]
[1]Universitas Indonesia, Depok, Indonesia, adib.prima@ui.ac.id
[2]Universitas Indonesia, Depok, Indonesia, rizal@cs.ui.ac.id

Corresponding Author: adib.prima@ui.ac.id[1]

Abstract: Organisations across finance, healthcare, public administration, and manufacturing have made artificial intelligence a central part of their operations. Yet a practical problem remains unsolved: practitioners still lack systematic, evidence-based guidance on which data governance practices reduce the risks that AI deployment brings. The existing literature addresses AI governance frameworks and responsible AI principles, but at a level of generality that offers limited operational value. No previous systematic literature review has assembled an empirically validated catalogue of data governance practices oriented specifically toward AI risk management. This study addresses that gap directly. Drawing on peer-reviewed empirical research published between 2020 and 2026, we followed the Kitchenham and Charters protocol and PRISMA 2020 guidelines, searched six electronic databases, and arrived at a final corpus of 21 high-quality studies. From these, we extracted and classified 96 data governance practices across eight domains: Data Infrastructure and Architecture Governance (D1), Data Lineage, Traceability and Auditability (D2), Data Quality and Integrity Risk Management (D3), Algorithmic Accountability, Explainability and Bias Control (D4), Data Privacy and Security Governance (D5), Regulatory Compliance and Legal Alignment (D6), AI Risk Assessment and Governance Framework (D7), and Organizational and Socio-Technical AI Governance (D8). D1, D7, and D8 together account for 57.3% of all practices. D4 is the thinnest domain, with only 6 practices (6.3%), even as regulatory pressure around explainability intensifies. Three domains, D4, D6, and D7, have no direct counterparts in DAMA-DMBOK, confirming that AI governance introduces requirements that conventional data management frameworks were not designed to handle. The mean quality assessment score across included studies was 5.57 out of 6.0. This catalogue is, to our knowledge, the first of its kind grounded entirely in empirical evidence from real-world organisational settings.

Keyword: Data Governance, Artificial Intelligence, Risk Management, Systematic Literature Review, AI Governance, Responsible AI, Machine Learning.

INTRODUCTION

Few technologies have been adopted as quickly and as widely as artificial intelligence. Financial institutions, healthcare providers, government agencies, and manufacturers have invested heavily in AI-based systems designed to automate decisions, streamline processes, and derive insight from large datasets (Joshi et al., 2025; Lestari et al., 2025; Chaturvedi, 2025). The pace of adoption has been remarkable. So, too, has been the pace at which the risks have come into view. Organisations deploying AI now regularly encounter problems they did not fully anticipate: models that produce discriminatory outputs, privacy violations arising from training data handling, failures of regulatory compliance, and persistent accountability gaps (Mandal et al., 2025; Pasam et al., 2025). These are not edge cases. They reflect structural governance failures.

Running through all of these risks is a single root problem: *data governance.* Whether an AI system produces reliable, fair, and trustworthy outputs depends directly on how the underlying data is managed across its lifecycle, specifically its quality, lineage, security, and ethical handling (Jacobs et al., 2025). When data governance is weak or absent, the consequences compound. Biased training data produces discriminatory outputs. Untracked lineage makes auditing impossible. Inadequate access controls expose sensitive information. And without quality monitoring in place, model degradation goes undetected until damage has already occurred (Fu & Liang, 2025).

That these challenges are widely recognised has not made them easier to resolve. Empirical studies consistently document the same shortfalls: AI systems trained on unrepresentative data, absent provenance records, inadequate consent mechanisms for training datasets, and models deployed without any interpretability mechanism (Scott et al., 2026). Practitioners know what good governance looks like in theory; closing the gap between that theory and operational practice is another matter. The regulatory environment further complicates things. The EU AI Act (European Parliament & Council, 2024), GDPR, and a growing body of national AI governance instruments are imposing specific new data governance obligations on organisations that operate AI systems obligations whose practical implications are often unclear.

Prior secondary reviews have addressed AI adoption in organisations (Raber et al., 2023), governance principles (Deibler et al., 2024), responsible AI (Laux et al., 2022; Alzahrani et al., 2025), and AI risk management (Kurniawan et al., 2023; Zhu et al., 2025). These contributions are useful, but they share a common limitation: none focuses specifically on *data governance practices* as operational tools for managing AI risk. Several blend theoretical proposals with empirical findings without distinguishing between them. And most draw primarily on pre-2020 literature, missing the regulatory developments that have since fundamentally reshaped what organisations are expected to do. This is the gap we address.

Our response is a systematic literature review (SLR) designed to identify, extract, and synthesise empirically validated data governance practices for AI risk management. We make three main contributions. First, we offer a structured catalogue of 96 practices grouped into eight implementation domains, all drawn from real-world empirical sources. Second, we identify three governance domains (D4, D6, and D7) with no equivalent in established data management frameworks, suggesting AI governance is not simply an extension of traditional data governance. Third, we conduct a research gap analysis that reveals where the empirical literature is thinnest, most notably in Algorithmic Accountability and Regulatory Compliance.

The paper proceeds as follows. Section II provides background on data governance and AI risk management. Section III surveys prior secondary studies. Section IV details the SLR methodology. Section V presents and discusses findings. Section VI draws conclusions and outlines future directions.

METHOD

We designed this SLR following Kitchenham and Charters (Kitchenham & Charters, 2007) and reported it according to the PRISMA 2020 statement (Page et al., 2021). Both choices were deliberate: these protocols are widely adopted in software engineering and computing research, and they provide a systematic basis for minimising selection bias and promoting transparent evidence synthesis. The review proceeded in three phases: planning, conducting, and reporting.

Research Questions

Three research questions shaped the inquiry:

RQ1: What data governance practices for risk management in AI implementation are reported in empirical literature?

RQ2: How are data governance practices for AI risk management distributed across domains, and what patterns emerge across organizational and industry contexts?

RQ3: What gaps exist in current empirical research on data governance practices for AI risk management?

Search Strategy

We limited the search to peer-reviewed literature published between 2020 and 2026. This window was selected deliberately: it captures the period of most significant regulatory and governance development in AI, from the emergence of the EU AI Act and ISO 42001 through the recent surge in generative AI governance activity. Literature predating 2020 would reflect a different context, one with fewer regulatory anchors. We queried six databases: IEEE Xplore, ProQuest, ScienceDirect, Scopus, Springer Nature Link, and Taylor & Francis Online. The multi-database approach reduces publisher bias and ensures adequate coverage across technical, interdisciplinary, and applied research streams.

We ran all database queries on 22 April 2026. A Boolean search string was developed and then adapted for each database's syntax and field operators. The core logic is as follows:

("data governance" OR "data management" OR "data stewardship" OR "data quality management" OR "data lineage") AND ("artificial intelligence" OR "machine learning" OR "AI system" OR "AI implementation" OR "AI-driven") AND ("risk management" OR "risk mitigation" OR "risk assessment" OR "risk control" OR "risk identification") AND (practice OR approach OR method OR "case study" OR implementation OR framework)

We validated the search string by checking that three known relevant papers, (Jacobs et al., 2025; Huang et al., 2026; Wang et al., 2026), were retrieved across their respective databases. All three were found. This gave us reasonable confidence in recall, though we acknowledge that formal precision and recall metrics were not computed.

Table 1. Final Search String per Database

Database	Boolean Search String
IEEE Xplore	("All Metadata":"data governance" OR "data management" OR "data stewardship" OR "data quality management" OR "data lineage") AND ("All Metadata":"artificial intelligence" OR "machine learning" OR "AI system" OR "AI implementation" OR "AI-driven") AND ("All Metadata":"risk management" OR "risk mitigation" OR "risk assessment" OR "risk control" OR "risk identification") AND ("All Metadata":practice OR approach OR method OR "case study" OR implementation OR framework)
ProQuest	(TI/AB/SU("data governance" OR "data management" OR "data stewardship")) AND (TI/AB/SU("artificial intelligence" OR "machine learning" OR "AI system")) AND (TI/AB/SU("risk management" OR "risk mitigation" OR "risk assessment")) AND (TI/AB/SU(practice OR approach OR method OR "case study" OR implementation OR framework))

Database	Boolean Search String
ScienceDirect	"data governance" AND "artificial intelligence" AND "risk management" AND "case study" [Filter: Research articles]
Scopus	TITLE-ABS-KEY(("data governance" OR "data management" OR "data stewardship" OR "data quality management" OR "data lineage") AND ("artificial intelligence" OR "machine learning" OR "AI system" OR "AI implementation" OR "AI-driven") AND ("risk management" OR "risk mitigation" OR "risk assessment" OR "risk control" OR "risk identification") AND (practice OR approach OR method OR "case study" OR implementation OR framework))
Springer Nature Link	("data governance" OR "data management") AND ("artificial intelligence" OR "machine learning") AND ("risk management" OR "risk assessment") AND (implementation OR "case study") [Filter: Articles, English, Computer Science]
Taylor & Francis	("data governance" OR "data management" OR "data stewardship") AND ("artificial intelligence" OR "machine learning" OR "AI system") AND ("risk management" OR "risk mitigation" OR "risk assessment") AND (practice OR approach OR method OR "case study" OR implementation OR framework)

Inclusion and Exclusion Criteria

Before conducting the search, we established a structured set of inclusion and exclusion criteria. The criteria were applied sequentially: objective bibliographic filters first, then substantive content evaluation. The rationale was to separate mechanical screening from interpretive judgement. Table III presents the complete set.

Table 2. Study Selection Guidelines

Code	Inclusion Criteria (IC)
IC.1	Not a duplicate publication
IC.2	Published between 2020 and 2026
IC.3	Written in English
IC.4	Peer-reviewed publications (journal articles, conference papers)
IC.5	Primary focus on data governance practices, data management practices, or data-related risk management approaches applied in the context of AI system implementation, AI model development, or AI-driven systems
IC.6	Full text accessible through institutional or open access
IC.7	Papers must report on the practical application, evaluation, or observed results of data governance practices for risk management in AI implementation, validated through at least one of the following: case study, expert interview, practitioner survey, field experiment, or action research in an industrial or real-world organizational context
IC.8	Papers must address at least one identifiable data governance practice in relation to AI-specific risks
Code	**Exclusion Criteria (EC)**
EC.1	Duplicate publications (retain most complete or recent version)
EC.2	Published outside 2020–2026 timeframe, or erratum, corrigendum, retraction, or correction
EC.3	Not written in English
EC.4	Grey literature, books, book chapters, theses, dissertations, technical reports, posters, workshops, editorials, non-research content, short papers (<4 pages), or not peer-reviewed papers
EC.5	Papers where Data Governance and AI Implementation are not substantively connected—including papers that discuss data governance in non-AI contexts, or AI systems where data governance is mentioned only tangentially

EC.6	Full text not accessible through institutional or open access
EC.7	Papers that are purely theoretical, conceptual, or solution proposals (new models, frameworks, or methodologies) that lack documented application or validation in real-world projects
EC.8	Papers that focus exclusively on technical AI model aspects without addressing data governance or organizational risk management dimensions
EC.9	Papers that address data governance exclusively in non-AI digital systems without extension to AI implementation context

Study Selection

The database search returned 2,236 records: IEEE Xplore (138), ProQuest (59), ScienceDirect (368), Scopus (278), Springer Nature Link (507), Taylor & Francis (886). After removing 692 duplicates, 1,544 unique records proceeded to screening. Objective filtering excluded 1,321 records: 7 fell outside the 2020–2026 window, 4 were not in English, 89 were not peer-reviewed, and 1,221 lacked a substantive DG-AI connection. Of 223 papers sought for full-text retrieval, 77 were inaccessible. At the eligibility stage, 125 of the remaining 146 were excluded: 94 for lacking real-world validation, 18 for focusing on AI without engaging data governance or organisational risk, 12 for treating DG only in non-AI contexts, and one for failing quality assessment. The PRISMA 2020 flow appears as Figure 1. The final corpus comprised 21 studies.

Quality Assessment

We applied quality assessment to all 36 papers reaching the eligibility stage. Each was scored against six questions (QA1–QA6), rated 0 (not met), 0.5 (partially met), or 1.0 (fully met), with a maximum of 6.0. The questions asked: (QA1) whether objectives and methodology are clearly described; (QA2) whether data governance practices are identified within an AI implementation context; (QA3) whether the link between DG practices and AI risk is explicit; (QA4) whether findings draw on empirical evidence from real-world settings; (QA5) whether organisational or sector context is described adequately; and (QA6) whether limitations are acknowledged and conclusions are supported by the evidence.

Papers scoring below 4.0 were excluded. One paper scored 3.5 and was removed. The remaining 21 studies form the final evidence base. Their mean QA score was 5.57 out of 6.0, reflecting a high level of methodological rigour across the corpus.

Data Extraction and Synthesis

For each of the 21 included studies, we extracted data using a structured per-practice instrument maintained in digital form. We recorded paper type, research method, industry sector, organisational scale, AI system type, lifecycle phase, QA score, and individual practices. A key decision was to extract practices at a granular level: each independently implementable governance action was recorded as a separate entry rather than being aggregated at the mechanism or framework level. A paper describing a governance approach with five distinct components therefore contributed five separate entries to the catalogue, yielding a richer and more practically usable inventory.

Extracted practices were then grouped into implementation domains through iterative qualitative classification. The first author made an initial domain assignment for each practice based on functional characteristics and thematic context. The co-author reviewed the structure for consistency, mutual exclusivity, and completeness. Disagreements were resolved through discussion. The final eight-domain taxonomy was validated against five reference frameworks: NIST AI RMF (National Institute of Standards and Technology, 2023), ISO/IEC 42001:2023 (International Organization for Standardization, 2023), DAMA-DMBOK (DAMA International, 2017), EU AI Act (European Parliament & Council, 2024), and FAIR

Principles (Wilkinson et al., 2016). We note that formal inter-rater reliability measurement was not conducted, which we acknowledge as a limitation.

RESULTS AND DISCUSSION

Statistics

The 21 studies in the final corpus span 2020 to 2026. The temporal distribution is uneven: one study dates from 2021, one from 2024, thirteen from 2025, and six from 2026. The sharp concentration in 2025 reflects the surge in AI governance research that followed the enactment of the EU AI Act in 2024 and the publication of ISO 42001 in 2023. The research community appears to have responded quickly to those regulatory signals.

The corpus comprises 12 journal articles (57.1%) and 9 conference papers (42.9%). The mean QA score of 5.57/6.0 is consistent across both publication types, suggesting methodological quality was not concentrated in one venue. Table IV shows the distribution by publisher. Elsevier contributes the largest group at 9 papers, spanning Technology in Society, Information Resources Management Journal, Computers & Industrial Engineering, and Government Information Quarterly, among others. IEEE contributes 5 papers. The remaining 7 are distributed across PLOS, Frontiers, ACM, IGI Global, IOS Press, ERDev, and MDPI.

Table 3. Publication Venues

Publisher	Publication Venues	No. of Papers
Elsevier	Technology in Society (2), Results in Engineering (1), IJIM (1), Computers & Industrial Engineering (1), Government Information Quarterly (1), Project Leadership & Society (1), IFAC PapersOnLine (1), Procedia Computer Science (1)	9
IEEE	IEEE MultiMedia (1), IEEE ISTAS (1), IEEE ICIMCIS (1), IEEE CogMI (1), IEEE ICCTDC (1)	5
PLOS	PLOS Digital Health (1)	1
Frontiers	Frontiers in Virtual Reality (1)	1
ACM	DMIT 2025 (1)	1
IGI Global	Information Resources Management Journal (1)	1
IOS Press	MEDINFO 2025 (1)	1
ERDev / Latvia	Engineering for Rural Development (1)	1
MDPI	Buildings (1)	1
Total		21

Overview of Selected Studies (RQ1)

Analysis of the 21 included studies produced 96 data governance practices, enumerated in Table V. All 96 practices come from studies with documented real-world validation: 10 used case study methods, 5 combined expert surveys with direct field implementations, 2 used semi-structured practitioner interviews, 1 applied AHP analysis to practitioner survey data, and 3 used controlled experiments or benchmarking. The catalogue therefore reflects what organisations are actually doing, rather than what researchers or framework authors think they should do.

Table 4. Practices per Paper

Study ID	Author	Year	Practices Identified	Domain	Sec.
S01	Joshi, H., Hassani, S., Gandhi, D., and Hartman, L.	2025	Risk-Based AI Tiering and Classification System	D7	D4

Study ID	Author	Year	Practices Identified	Domain	Sec.
			Internal Sandbox Environment for Safe LLM Experimentation	D7	D5
			Shadow AI Detection and Governance Control	D7	D5
			Data Minimization Governance for GenAI Training Data	D5	D6
			Continuous GenAI Model Drift Prevention and Monitoring	D7	—
			GenAI Lifecycle Governance Checkpoint Embedding	D8	D7
			Automated Bias Detection using Responsible AI Tools	D4	—
S02	Lestari, M., Wijaya, A.F., Sari, M.K., and Leander, L.K.	2025	Secure MLOps Pipeline with Dataset Versioning and Lineage Enforcement	D1	D2
			Multi-Standard AI Governance Framework Integration (ISO 42001, NIST AI RMF, COBIT 2019)	D7	D6
			Zero Trust Architecture Implementation for AI Data Access	D5	D1
			Ambidextrous AI Governance Maturity Assessment	D8	D7
			Explainability Dashboard and Automated Risk Monitoring	D4	D7
S03	Chaturvedi, B.	2025	NLP-Based Intelligent Data Topic Routing in Kafka Streaming	D1	—
			ML-Based Real-Time Data Quality Assessment in Flink Streaming	D3	—
			Schema Evolution Detection and Automatic Pipeline Updates	D1	D3
			SHAP and LIME Explainability Integration for Regulatory Compliance	D4	D6
			BERT-Based Automated Regulatory Compliance Mapping	D6	D1
			Immutable Automated Audit Trail Generation for All Data Flows	D2	D6
			Real-Time NLP Regulatory Rule Extraction and Policy Updates	D6	—
			Reinforcement Learning-Based Autonomous Data Pipeline Routing	D1	D7
S04	Mandal, A., Ramanayake, R., O'Neill, O., et al.	2025	Use-Case Specific LLM Risk Tiering and Governance Categorization	D7	—
			Comprehensive LLM Pre-Deployment Evaluation with Sector-Specific Benchmarks	D7	D4
			Fairness Constraints and Bias	D4	—

Study ID	Author	Year	Practices Identified	Domain	Sec.
			Mitigation for LLM Banking Applications		
			Human-in-the-Loop Oversight Framework for LLM Banking Decisions	D8	D7
			LLM Guardrails Implementation for Safety and Compliance	D7	D5
			AI Ethics Officer Designation for LLM Oversight	D8	—
			Detailed LLM Usage Documentation and Audit Trail Management	D2	D6
S05	Pasam, V.R., Krishnan, B., Charla, R.R., Somayajula, R., and Veerapaneni, S.M.	2025	AI-Augmented ETL with Isolation Forests and KNN Imputation	D1	D3
			Data Lineage Recording via Neo4j Graph Database	D2	D1
			Automated Data Quality Scoring for Every Pipeline Batch	D3	D2
			Real-Time Data Pipeline Orchestration Governance	D1	—
S06	Jacobs, J.J.L., Beekers, I., Verkouter, I., et al.	2025	FAIR Data Dictionary Implementation for Multi-Center Clinical AI	D2	D1
			Attribute-Based Access Control (ABAC) for Medical Data	D5	—
			Synchronized Multi-Center Clinical Template Forms	D1	D3
			Privacy-by-Design Architecture for Precision Medicine AI	D5	D6
			Cross-Institutional Research Data Governance Framework	D6	D5
S07	Fu, Y. and Liang, D.	2025	Four-Layer Data Governance Architecture for Enterprise AI	D1	—
			Standardized Enterprise Data Naming and Dictionary System	D1	D3
			ML-Based Intelligent Decision System for Risk Assessment	D7	—
			Microservice-Based Independent Service Unit Deployment	D1	—
S08	Scott, J.A., Bagade, A., and Choudhary, B.	2026	IRB Tiered Assessment Framework for AI and XR Technologies	D8	D7
			Data Usage Agreement Criteria for AI Research Systems	D6	D5
			AI Inference Type Documentation Requirement for Research Protocols	D7	D4
			Enhanced Anonymization Protocol for High-Dimensional AI Data	D5	—

Study ID	Author	Year	Practices Identified	Domain	Sec.
S09	Iders-Bankovs, M., Politika, V., Pundure, J., Jarvis, M., and Ziemelis, M.	2025	Predictive Parity Rate and Demographic Parity Metrics for AI Procurement	D4	—
			Explainable AI Implementation for Transparent Public Procurement	D4	D6
			Continuous AI System Monitoring for Procurement Fairness	D7	D4
			Advanced Data Management System Development for AI Procurement	D1	—
			AI Literacy Enhancement Program for Procurement Staff	D8	—
S10	Yuan, J.	2025	Automated ETL Pipeline with Data Cleansing and Multi-Source Integration	D1	D3
			Four-Layer Big Data Architecture for Financial Risk Management	D1	—
			NLP and Random Forest Risk Scoring Integration	D7	D1
S11	Freeman, S. et al.	2025	Risk Tiering for Healthcare AI Committee Approval	D7	—
			Federated Data Schema for Sensitive Healthcare Data	D5	D1
			Centralized AI Governance Board for Healthcare Organizations	D8	D7
			AI Literacy Program for Healthcare Staff	D8	—
S12	Almadhoob, H.	2026	Permissioned Blockchain for Government Approval Audit Trail	D2	D1
			AI Validation Combined with Smart Contract Automation	D7	D1
			Centralized AI Verification Portal as Single Point of Truth	D1	D8
S13	Baharmand, H.	2025	Bespoke GenAI Instance Design for Vulnerable Population Data	D5	D7
			Governance Safeguards for Humanitarian GenAI Data Anti-Leakage	D7	D5
			Cross-Sector Data Governance Collaboration Framework	D8	D6
			AI Literacy Program for Humanitarian Data Workers	D8	—
S14	Mapupu, T. and Nwaila, G.T.	2026	IIoT Sensor Real-Time Anomaly Validation at Data Ingestion	D1	D3
			AES-256 Encrypted Data Transmission for Industrial Sensors	D5	D1
			Hybrid Cloud-Local Architecture for Metal Accounting Data Governance	D1	D5

Study ID	Author	Year	Practices Identified	Domain	Sec.
			AMIRA P754 Code of Practice Compliance Framework	D6	D3
			Multi-Stage QA/QC Data Quality Control Protocol	D3	D2
S15	Huang, X., Kou, T., and Zhou, Q.	2026	Authorization Tracking at AI Data Acquisition Phase	D2	D6
			Ethics-Based Quality Control Checkpoint at AI Model Evaluation	D3	D4
			Secure Deletion Protocol for AI Data Termination Phase	D5	D6
			FAHP-Based Ethics Indicator Weighting for AI Governance Prioritization	D8	D7
			Vertical Feedback Mechanism Embedding in AI Governance Lifecycle	D8	—
S16	Wang, X., Zhong, W., Huang, K., and Liang, B.	2026	Split-Model Data Security Strategy for Sensitive Organizational Data	D5	D7
			Comprehensive Access Permission Logging for LLM Systems	D2	D5
			Tiered Rollout Strategy for Organizational LLM Implementation	D8	D7
			Agile Modular Architecture Governance for LLM Integration	D1	D8
			Socio-Technical Contradiction Analysis for LLM Governance Design	D8	—
S17	Grunt, M., Blazejewski, A., Pecolt, S., and Krolikowski, T.	2025	Data Aggregation Pipeline from Heterogeneous Legacy Security Systems	D1	D3
			ML-Based Predictive Anomaly Detection for Cyber-Physical Security	D7	D1
			10-Point Structured AI/ML Implementation Plan for Security Governance	D8	D7
S18	Perdana, A., Arifin, S., and Quadrianto, N.	2025	Centralized Data Classification Governance under Central Bank Authority	D5	D6
			ISO Cybersecurity Standards Implementation for AI-Driven Banking Systems	D6	D5
			GELSI Framework Implementation for Algorithmic Financial Governance	D8	D7
			Ethical Guidelines Prioritization for Financial Algorithmic Systems	D8	D4
S19	Pesqueira, A., de Bem Machado, A., Bolog, S.,	2024	EU AI Act Alignment Protocol for Pharmaceutical AI Systems	D6	D7

Study ID	Author	Year	Practices Identified	Domain	Sec.
	Pereira, R., and Sousa, M.J.				
			GDPR Compliance Protocol in AI-Assisted Pharmaceutical Procurement	D6	D5
			Expert Consensus Governance via Delphi Method for AI Standards	D8	D7
S20	Vetrò, A., Torchiano, M., and Mecati, M.	2021	Imbalance Ratio as Formal Risk Indicator under ISO 31000	D3	D4
			Gini and Simpson Index Application for Discrimination Risk Assessment	D3	D4
			ISO 31000 Risk-Based Framework for ADM Data Quality Governance	D7	D3
S21	Jou, T.S., Maaz, Z.N., Hanid, M., et al.	2026	Traceability Prioritization via Distributed Blockchain Ledger	D2	D1
			Interoperability Implementation via BIM/IoT Integration in CDE	D1	D3
			Phased Data Governance Implementation for Construction AI	D8	D1
			System-Dependent Data Quality Dimension Prioritization Framework	D3	D2
			Data Provenance Verification Framework for AI Construction Decisions	D2	D7
TOTAL	**96 practices extracted from 21 empirically validated studies**				

Domain Mapping Summary (RQ2)

Turning to RQ2: the 96 practices were classified into eight implementation domains. Table 5 presents the full domain mapping, including practice counts, source studies, distributional shares, and cross-study coverage percentages, with illustrative examples for each domain.

Table 5. Domain Mapping Summary

Domain	Domain Name	Total Practices	Studies (Study IDs)	Dist. %	Cov. %	Example Practices
D1	Data Infrastructure & Architecture Governance	19	S02, S03, S05, S06, S07, S09, S10, S12, S14, S16, S17, S21	19.8%	57.1%	Secure MLOps Pipeline with Dataset Versioning and Lineage Enforcement [S02]; Four-Layer Data Governance Architecture for Enterprise AI [S07]; NLP-Based Intelligent Data Topic Routing in Kafka Streaming [S03]
D2	Data Lineage, Traceability & Auditability	9	S03, S04, S05, S06, S12, S15, S16, S21	9.4%	38.1%	Data Lineage Recording via Neo4j Graph Database [S05]; Permissioned Blockchain for

Domain	Domain Name	Total Practices	Studies (Study IDs)	Dist. %	Cov. %	Example Practices
						Government Approval Audit Trail [S12]; FAIR Data Dictionary Implementation for Multi-Center Clinical AI [S06]
D3	Data Quality & Integrity Risk Management	7	S03, S05, S14, S15, S20, S21	7.3%	28.6%	ML-Based Real-Time Data Quality Assessment in Flink Streaming [S03]; Imbalance Ratio as Formal Risk Indicator under ISO 31000 [S20]; Multi-Stage QA/QC Data Quality Control Protocol [S14]
D4	Algorithmic Accountability, Explainability & Bias Control	6	S01, S02, S03, S04, S09	6.3%	23.8%	SHAP and LIME Explainability Integration for Regulatory Compliance [S03]; Fairness Constraints and Bias Mitigation for LLM Banking Applications [S04]; Predictive Parity Rate and Demographic Parity Metrics [S09]
D5	Data Privacy & Security Governance	11	S01, S02, S06, S08, S11, S13, S14, S15, S16, S18	11.5%	47.6%	Split-Model Data Security Strategy for Sensitive Organizational Data [S16]; Privacy-by-Design Architecture for Precision Medicine AI [S06]; Enhanced Anonymization Protocol for High-Dimensional AI Data [S08]
D6	Regulatory Compliance & Legal Alignment	8	S03, S06, S08, S14, S18, S19	8.3%	28.6%	EU AI Act Alignment Protocol for Pharmaceutical AI Systems [S19]; BERT-Based Automated Regulatory Compliance Mapping [S03]; AMIRA P754 Code of Practice Compliance Framework [S14]

Domain	Domain Name	Total Practices	Studies (Study IDs)	Dist. %	Cov. %	Example Practices
D7	AI Risk Assessment & Governance Framework	17	S01, S02, S04, S07, S08, S09, S10, S11, S12, S13, S17, S20	17.7%	57.1%	Multi-Standard AI Governance Framework Integration — ISO 42001, NIST AI RMF, COBIT 2019 [S02]; Risk Tiering for Healthcare AI Committee Approval [S11]; LLM Guardrails Implementation for Safety and Compliance [S04]
D8	Organizational & Socio-Technical AI Governance	19	S01, S02, S04, S08, S09, S11, S13, S15, S16, S17, S18, S19, S21	19.8%	61.9%	GenAI Lifecycle Governance Checkpoint Embedding [S01]; Tiered Rollout Strategy for Organizational LLM Implementation [S16]; FAHP-Based Ethics Indicator Weighting for AI Governance Prioritization [S15]
TOTAL	**Eight empirically-grounded data governance domains**	**96**		**100%**	—	—

Two domains tied for the highest practice count: Data Infrastructure and Architecture Governance (D1) and Organizational and Socio-Technical AI Governance (D8), each contributing 19 practices (19.8%), sourced from 12 and 13 studies respectively. D1's prominence is broadly expected; data infrastructure is the layer on which all other governance depends. Without reliable pipelines, enforced schema governance, and architectural integrity, the other seven domains become difficult to implement. D8's equal prominence is theoretically more interesting. It confirms a point the NIST AI RMF implies but does not always make explicit: technical data governance cannot function without commensurate organisational structures, human capacity development, and accountability mechanisms.

AI Risk Assessment and Governance Framework (D7) ranked second, with 17 practices (17.7%) from 12 studies (57.1% coverage). This domain captures specifically risk-oriented practices: risk tiering, scoring, continuous monitoring. Worth noting is that D7 has no equivalent in DAMA-DMBOK, which reflects how AI governance introduces risk dimensions that traditional data management frameworks were simply not designed to address. Data Privacy and Security Governance (D5) contributed 11 practices (11.5%) from 10 studies (47.6% coverage). The relatively high cross-study coverage here reflects consistent regulatory pressure around data privacy, driven primarily by GDPR but increasingly by AI-specific instruments.

The least populated domain is Algorithmic Accountability, Explainability and Bias Control (D4), with only 6 practices (6.3%) from 5 studies (23.8% coverage). This is the most concerning finding in the entire analysis. The EU AI Act directly mandates accountability, transparency, and explainability mechanisms in Articles 13, 14, and 27. Yet the empirical evidence documenting actual implementation of such mechanisms is almost entirely absent. The gap between what regulation requires and what organisations have documented as practice is striking, and it provides a direct response to RQ3.

Paper-Domain Matrix (RQ2)

Table 6 shows the paper-domain matrix, mapping how individual studies distribute their practices across domains. Most studies (14 of 21) contributed to two or three domains. Four were confined to a single domain, while three engaged four or more. The broadest single contribution came from S03 (Chaturvedi), which spanned five domains, reflecting the cross-cutting nature of a cognitive data architecture designed for financial services AI compliance.

Table 6. Paper-Domain Matrix

Study	Authors	D1	D2	D3	D4	D5	D6	D7	D8	Total
S01	Joshi et al. (2025)	—	—	—	1	1	—	4	1	7
S02	Lestari et al. (2025)	1	—	—	1	1	—	1	1	5
S03	Chaturvedi (2025)	3	1	1	1	—	2	—	—	8
S04	Mandal et al. (2025)	—	1	—	1	—	—	3	2	7
S05	Pasam et al. (2025)	2	1	1	—	—	—	—	—	4
S06	Jacobs et al. (2025)	1	1	—	—	2	1	—	—	5
S07	Fu and Liang (2025)	3	—	—	—	—	—	1	—	4
S08	Scott et al. (2026)	—	—	—	—	1	1	1	1	4
S09	Iders-Bankovs et al. (2025)	1	—	—	2	—	—	1	1	5
S10	Yuan (2025)	2	—	—	—	—	—	1	—	3
S11	Freeman et al. (2025)	—	—	—	—	1	—	1	2	4
S12	Almadhoob (2026)	1	1	—	—	—	—	1	—	3
S13	Baharmand (2025)	—	—	—	—	1	—	1	2	4
S14	Mapupu and Nwaila (2026)	2	—	1	—	1	1	—	—	5
S15	Huang et al. (2026)	—	1	1	—	1	—	—	2	5
S16	Wang et al. (2026)	1	1	—	—	1	—	—	2	5
S17	Grunt et al. (2025)	1	—	—	—	—	—	1	1	3
S18	Perdana et al. (2025)	—	—	—	—	1	1	—	2	4
S19	Pesqueira et al. (2024)	—	—	—	—	—	2	—	1	3
S20	Vetrò et al. (2021)	—	—	2	—	—	—	1	—	3
S21	Jou et al. (2026)	1	2	1	—	—	—	—	1	5
TOTAL	**All 21 studies**	**19**	**9**	**7**	**6**	**11**	**8**	**17**	**19**	**96**

A clear co-occurrence pattern emerges from the matrix: studies engaging D1 (Data Infrastructure) reliably also engage D7 (AI Risk Assessment) and D8 (Organisational Governance). This is consistent with the NIST AI RMF's integrated logic, which treats technical infrastructure governance as inseparable from organisational accountability structures and risk assessment processes. Financial sector studies (S03, S07, S10) cluster heavily in D1, reflecting the centrality of pipeline integrity for financial AI risk. Cross-sector and organisational behaviour studies (S01, S04, S15, S16) concentrate in D7 and D8, capturing the governance and sociotechnical dimensions that characterise GenAI adoption.

Research Gap Analysis (RQ3)

We examined research gaps along two dimensions: domain coverage rates (Table 6) and the temporal distribution of practices across years (Table 7). Coverage, defined as the proportion of the 21 studies engaging each domain, varies markedly across the taxonomy.

D4 is the clearest gap, with coverage of only 23.8% from 5 studies. Given the regulatory attention this area receives under the EU AI Act (Articles 13, 27, 53) and a growing body of national frameworks, the near-absence of empirical work documenting actual implementation of accountability, explainability, and bias control mechanisms is notable. Regulatory demands and research supply appear substantially misaligned here. D3 (Data Quality) and D6 (Regulatory Compliance) also trail at 28.6% coverage each, representing meaningful underrepresentation for two domains whose importance is broadly acknowledged.

Table 7. Temporal Distribution of Practices per Domain

Domain	2021	2024	2025	2026	Total
D1 — Data Infrastructure	—	—	14	5	19
D2 — Data Lineage, Traceability	—	—	4	5	9
D3 — Data Quality	2	—	2	3	7
D4 — Algorithmic Accountability, Explainability	—	—	6	—	6
D5 — Data Privacy	—	—	7	4	11
D6 — Regulatory Compliance	—	2	4	2	8
D7 — AI Risk Assessment	1	—	14	2	17
D8 — Organizational	—	1	12	6	19
TOTAL	**3**	**3**	**63**	**27**	**96**

The temporal data adds a further layer of nuance. D1 and D7 show consistent activity across all years, with a marked peak in 2025, where each produced 14 practices. This suggests data infrastructure and risk assessment became the central governance concerns during that year's AI adoption surge. D8 (Organisational Governance) shows the strongest growth specifically in 2026, with 6 practices appearing that year. This late surge likely reflects the crystallisation of governance board structures, AI ethics frameworks, and sociotechnical governance models as organisations began substantively responding to regulatory pressure.

The temporal profile of D4 deserves particular attention. All 6 of its practices cluster in 2025, with nothing from 2021, 2024, or 2026. Combined with its low overall coverage, this tells a clear story: empirical implementation of algorithmic accountability has not kept pace with either the regulatory discourse or the academic conversation. The EU AI Act centres transparency and explainability in its high-risk AI obligations, yet documented evidence of organisations actually implementing and validating XAI mechanisms in governance contexts remains sparse across the entire 2020–2026 literature window studied.

Cross-Domain Analysis and Synthesis

Beyond single-domain analysis, the matrix reveals something important about how governance operates in practice: domains do not function independently. The most consistent pairing across the 21 studies is D1 with D7 and D8. Studies addressing data infrastructure almost invariably also address AI risk assessment and organisational governance. This cluster appears across different sector contexts and is consistent with the NIST AI RMF's integrated logic. The implication for practitioners is concrete: effective data infrastructure governance for AI cannot be implemented without simultaneously building the risk assessment processes and organisational accountability structures that give that governance substance.

A second co-occurrence of theoretical significance links D3 (Data Quality) with D4 (Algorithmic Accountability). The most direct empirical evidence comes from S20 (Vetro et al., 2021), where Vetro et al. demonstrate that training data imbalance ratio, a standard data quality measure, directly predicts discriminatory outputs in automated decision-making systems. This is a causal mechanism, not merely a correlation. For practitioners, the implication is direct: algorithmic accountability work must begin with the data. Organisations that attempt to address model fairness and explainability without first establishing data quality governance are treating symptoms rather than causes.

Evidence strength is unevenly distributed. Practices in D1, D7, and D8 are corroborated across multiple independent studies in diverse organisational contexts, giving them relatively robust evidentiary standing. D4 practices, by contrast, come from just five studies, limiting their generalisability. Theoretically, the absence of DAMA-DMBOK counterparts for D4, D6, and D7 (DAMA International, 2017) is itself a significant finding: it confirms that AI implementation introduces governance requirements that conventional data management frameworks were never designed to handle. This is not a gap that can be filled by extending existing frameworks; it requires new ones.

CONCLUSION

This review set out to address three research questions about data governance practices for AI risk management. Drawing on 21 empirically grounded studies published between 2020 and 2026, we extracted and classified 96 implementation practices across eight domains. The overall finding is that data governance is not a single activity but a multi-dimensional enabler of responsible AI, spanning technical infrastructure and architecture (D1), lineage and traceability (D2), data quality (D3), algorithmic accountability (D4), privacy and security (D5), regulatory compliance (D6), risk assessment frameworks (D7), and organisational governance (D8). Several of these dimensions are interdependent in ways with direct practical implications for how governance investments should be sequenced.

On RQ1, this study provides the first comprehensive, empirically grounded catalogue of data governance practices for AI risk management. On RQ2, D1, D7, and D8 collectively account for 57.3% of all identified practices, confirming that data infrastructure, risk governance, and organisational dimensions dominate current practice. On RQ3, D4 (Algorithmic Accountability) is the clearest research gap: 6 practices (6.3%) from 5 studies (23.8% coverage), despite central regulatory importance. The requirement for empirical validation in IC.7 and IC.8 is what principally distinguishes this review from prior secondary studies.

REFERENCE

Almadhoob, H. (2026). Rethinking Bahrain's digital permit governance: A GenAI-enabled framework. *Project Leadership and Society*. Elsevier. https://doi.org/10.1016/j.plas.2026.100224

Alzahrani, A., et al. (2025). A systematic review of responsible artificial intelligence principles and practice. *Applied System Innovation, 8*(4), 97. https://doi.org/10.3390/asi8040097

Baharmand, H. (2025). Leveraging generative artificial intelligence to address data management challenges in humanitarian operations. *IFAC PapersOnLine*. Elsevier. https://doi.org/10.1016/j.ifacol.2025.09.250

Chaturvedi, B. (2025). Cognitive data architecture for financial services: A benchmark-driven framework for real-time, AI-enabled compliance and risk management. In *Proceedings of the 2025 IEEE 7th International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE. https://doi.org/10.1109/CogMI67134.2025.00030

DAMA International. (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd ed.). Technics Publications.

Deibler, W., et al. (2024). AI governance: A systematic literature review. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00653-w

European Parliament & Council. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union*.

Freeman, S., et al. (2025). Key considerations for governing safe and responsible use of AI in healthcare. In *MEDINFO 2025*. *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI251062

Fu, Y., & Liang, D. (2025). Data governance framework and intelligent decision-making system in enterprise technology strategy. In *Proceedings of the 2025 International Conference on Digital Management and Information Technology (DMIT 2025)*. ACM. https://doi.org/10.1145/3736426.3736504

Grunt, M., Blazejewski, A., Pecolt, S., & Krolikowski, T. (2025). AI in the digital transformation of technical protection systems: The SAFEGUARD framework and pilot study. *Procedia Computer Science (KES 2025)*. Elsevier. https://doi.org/10.1016/j.procs.2025.10.076

Huang, X., Kou, T., & Zhou, Q. (2026). Embedding AI ethics in the data lifecycle: A framework for enterprise AI governance. *Technology in Society*. Elsevier. https://doi.org/10.1016/j.techsoc.2026.103261

Iders-Bankovs, M., Politika, V., Pundure, J., Jarvis, M., & Ziemelis, M. (2025). Public procurement in the age of AI: Challenges and opportunities. *Engineering for Rural Development*. https://doi.org/10.22616/ERDev.2025.24.TF198

International Organization for Standardization. (2023). *ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system*. ISO.

Jacobs, J. J. L., Beekers, I., Verkouter, I., et al. (2025). A data management system for precision medicine. *PLOS Digital Health*. https://doi.org/10.1371/journal.pdig.0000464

Jobin, A., Ienca, M., & Vayena, E. (2022). Ethics of AI: A systematic literature review of principles and challenges. In *Proceedings of the FAccT 2022*. ACM. https://doi.org/10.1145/3530019.3531329

Joshi, H., Hassani, S., Gandhi, D., & Hartman, L. (2025). Approaches to responsible governance of GenAI in organizations. In *Proceedings of the 2025 IEEE International Symposium on Technology and Society (ISTAS)*. https://doi.org/10.1109/ISTAS65609.2025.1126965

Jou, T. S., Maaz, Z. N., Hanid, M., et al. (2026). Trustworthy AI in sustainable building projects: Prioritizing data quality for risk management decisions. *Buildings*. MDPI. https://doi.org/10.3390/buildings16071462

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Tech. Rep. EBSE 2007-001). Keele University.

Kurniawan, R., et al. (2023). Artificial intelligence risk identification: Challenges, impacts, and mitigation strategies. *International Journal of Electrical and Computer Engineering, 3*(2). https://doi.org/10.62146/ijecbe.v3i2.109

Laux, J., Wachter, S., & Mittelstadt, B. (2022). Characteristics and challenges in the industries towards responsible AI: A systematic literature review. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-022-09634-1

Lestari, M., Wijaya, A. F., Sari, M. K., & Leander, L. K. (2025). Artificial intelligence governance for innovation and resilient data security in academic institutions. In *Proceedings of the 2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. IEEE.

Mandal, A., Ramanayake, R., O'Neill, O., Flanagan, W., Kemal, N., Yekrangi, M., Chatbri, H., & Martin, C. (2025). Governing framework for the responsible democratization of large-language models in banking. *IEEE MultiMedia / Internet Computing*. https://doi.org/10.1109/MIC.2025.3621976

Mapupu, T., & Nwaila, G. T. (2026). Metal accounting data acquisition and management using real-time sensors and artificial intelligence. *Results in Engineering*. Elsevier. https://doi.org/10.1016/j.rineng.2026.110384

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce.

Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pasam, V. R., Krishnan, B., Charla, R. R., Somayajula, R., & Veerapaneni, S. M. (2025). AI-enhanced data engineering for complex workflows: Lessons from retail intelligence and financial risk models. In *Proceedings of the 2025 International Conference on Computing Technologies & Data Communication (ICCTDC)*. IEEE. https://doi.org/10.1109/ICCTDC64446.2025.11158176

Perdana, A., Arifin, S., & Quadrianto, N. (2025). Algorithmic trust and regulation: Governance, ethics, legal, and social implications blueprint for Indonesia's central banking. *Technology in Society*. Elsevier. https://doi.org/10.1016/j.techsoc.2025.102838

Pesqueira, A., de Bem Machado, A., Bolog, S., Pereira, R., & Sousa, M. J. (2024). Exploring the impact of EU tendering operations on future AI governance and standards in pharmaceuticals. *Computers & Industrial Engineering*. Elsevier. https://doi.org/10.1016/j.cie.2024.110655

Raber, D., Winter, R., & Wortmann, R. (2023). The implementation of artificial intelligence in organizations: A systematic literature review. *Information & Management, 60*(8), 103816. https://doi.org/10.1016/j.im.2023.103816

Scott, J. A., Bagade, A., & Choudhary, B. (2026). Immersive technologies and AI generate novel challenges to human subjects' protections protocols. *Frontiers in Virtual Reality*. https://doi.org/10.3389/frvir.2026.1674326

Vetro, A., Torchiano, M., & Mecati, M. (2021). A data quality approach to the identification of discrimination risk in automated decision-making systems. *Government Information Quarterly*. Elsevier. https://doi.org/10.1016/j.giq.2021.101619

Wang, X., Zhong, W., Huang, K., & Liang, B. (2026). High interest but low adoption: Navigating organizations' journey towards generative artificial intelligence implementation. *International Journal of Information Management*. Elsevier. https://doi.org/10.1016/j.ijinfomgt.2025.103009

Wilkinson, M. D., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

Yuan, J. (2025). A big data-driven information model for enterprise financial risk management: Model development and empirical validation. *Information Resources Management Journal*. IGI Global. https://doi.org/10.4018/IRMJ.396698

Zhu, Y., et al. (2025). Artificial intelligence in risk management within the realm of construction projects: A bibliometric analysis and systematic literature review. *Journal of Innovation & Knowledge*. https://doi.org/10.1016/j.jik.2025.100711