International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

# The Overlooked Key to AI Success: Why Clean, Reliable Data Outperforms Bigger Models

**Ali Azghar Hussain Syed Abbas**

Distinguished Solutions Data/AI Architec, Data/AI

**Abstract**
Organizations are seeking ever larger models and novel architectures, yet the long-term potential of AI lies in cleaner, better-managed data. This paper presents a data-first perspective, which argues that high-quality, context-rich, and well-documented data outperforms even the most advanced algorithms in terms of accuracy, robustness, explainability, and life-cycle economics. It discusses common defects—such as label noise, schema inconsistencies, missing data, and stale features—that lead to biased estimates, poor generalization, and expensive operations. Conversely, sustained investment in data quality, semantic standardization, and end-to-end lineage delivers better long-term performance without requiring ever-larger models.
The paper proposes a Data-First AI model that integrates data profiling, automated validation, relabeling, standardized semantics, and observability into the AI development process. Data quality dimensions—including accuracy, completeness, consistency, timeliness, representativeness, and provenance—are measured in relation to model failure modes and operational risks. The concept of smart-sized training data is also introduced, which optimizes signal-to-noise ratios, training costs, and iteration speed.
The proposed model is transferable to domains such as healthcare, smart infrastructure, manufacturing, and marketing. Versioned semantics in today's controlled systems and traceable provenance can streamline compliance and after-action analysis. Intrinsic data quality, model behavior, and business outcomes (KPIs, incident rates, time-to-recover) are evaluated using a three-tier measurement scheme. Overall, investments in data yield higher returns than additional cycles of architectural innovation and reduce both costs and risk.

## 1. Introduction

Over the past decade, the trajectory of artificial intelligence has often been portrayed as a race toward ever larger models and smarter architectures. In reality, however, even modest investments in cleaning, standardizing, and governing the data on which models learn and act tend to deliver more durable performance and business value. Strategy and management accounts emphasize that competitive advantage in AI relies less on exotic algorithms and more on a scalable capacity to gather, organize, and re-use high-quality data assets (Marr 2021), (Davenport 2018), (Davenport et al. 2019). Anthropocentric and human-centered perspectives likewise caution that, in the absence of provenance, transparency, and contextual awareness, algorithmic outputs can amplify inaccuracies and erode trust, regardless of how advanced the underlying models are (Gigerenzer 2022), (Hosanagar 2020), (Nowotny 2021), (Zweig 2022), (Marcus and Davis 2019).

The data floor in a production environment defines the best performance that any model can realistically achieve. Noise in labels, schema changes, duplicate records, missingness patterns, and complex interaction effects are translated into biased estimates, broken training–serving alignment, and unstable, near–zero-signal training dynamics. MLOps practices can further exacerbate these defects when they rely on ad hoc feature patches, frequent retraining, or purely post-hoc explainability work. Recent practice-focused research on AI-assisted data cleaning and continuous data quality improvement (DQI) suggests that teams can achieve step-change gains by systematizing profiling, validation, standardization, and feedback-driven data cleansing, without any changes to the model itself (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (America 2025). Streaming analytics and modern data infrastructure are most valuable on the platform side, where semantic consistency and freshness from ingestion through inference are enforced so that training and serving remain aligned (Zahra et al. 2024).

Another critical factor is governance: master data management (MDM), lineage, and audit trails provide organizations with shared semantics and traceability. This foundation reduces risk, accelerates delivery, and simplifies regulatory scrutiny when compared to heterogeneous data landscapes and opaque algorithmic safety mechanisms, which are especially prevalent in domains such as healthcare and critical infrastructure (Chen and Decary 2020), (Baker and Xiang 2023), (Şerban and Lytras 2020). Ethical and legal scholarship supports this argument: credible AI depends on verifiable data sources, versioned transformations, and responsible stewardship, rather than ex-post model explanations alone (Lobel 2023), (Floridi 2023), (Chander et al. 2025).

Another persistent myth is that more data is always better. When label noise is the dominant factor, large volumes of false positives and false negatives can entrench spurious correlations and drive up costs. Recent work shows that so-called smart-sizing of the training set—focused growth when signal is sparse and targeted shrinkage when noise is high—tends to be more cost-effective, more accurate, and more robust than simply scaling up model size (Cook and Klawa 2025), (Alves et al. 2021). In other words, ever-larger algorithms alone will not make AI reliably trustworthy; the decisive leverage lies in higher-quality data, as evidenced by recent publications and practical experience in DQI, MDM, reputable-AI (Panwar 2024), (Thirunagalingam 2024) and related fields (America 2025), (Marr 2021), (Zahra et al. 2024), (Davenport 2018), (Chander et al. 2025).

***Data Quality Dimensions, Usual Defects and Downstream Impact***

| **Dimension (what to measure)** | **Common Defect Examples** | **Model-Level Impact** | **Ops / Compliance Impact** |
|---|---|---|---|
| Accuracy (truthfulness vs. ground truth) | Mislabels, unit mix-ups, transcription errors | Bias; degraded precision / recall; unstable convergence | Rework; incident tickets; audit findings |
| Completeness (expected fields present) | Systematic missingness; sparse critical fields | Selection bias; brittle inference paths | Manual patching; slowed releases |
| Consistency (schemas / semantics aligned) | Conflicting codes/IDs; divergent definitions | Feature leakage; poor generalization; drift | Cross-team disputes; duplicated logic |
| Timeliness / Freshness (SLA adherence) | Stale features; lagging updates | Drift; delayed reactions; degraded UX | On-call spikes; customer issues |
| Representative ness (coverage of realworld | Skewed sampling; domain gaps | OOD failure; fairness issues | Remediation cost; reputational risk |

| Dimension (what to measure) | Common Defect Examples | Model-Level Impact | Ops / Compliance Impact |
|---|---|---|---|
| variation) | | | |
| Provenance Lineage (traceable origin and transforms) | Unversioned joins; opaque ETL | Unexplained regressions; weak XAI | Compliance burden; change-approval delays |

A combination of the above dimensions yields a practical diagnostic grid: intrinsic information quality (e.g., schema conformance, freshness SLAs, inter-rater label accuracy) and model behavior (calibration, memory-of-retrain stickiness, out-of-distribution sensitivity) can be measured and tracked, and these indicators can be related to work outputs (incident rates, time-to-recover, compute spend per performance unit). Such traceability enables leaders to attribute gains to specific data interventions, such as standardizing high-leverage fields with MDM, eliminating duplicate entity keys, or renaming ambiguous classes, rather than relying on gut instinct about whether an incidental model change caused an improvement.
In medical care, versioned changes and provenance facilitate clinical certification and post hoc inspection, and increase safety and reliability at the point of care (Chen and Decary 2020), (Baker and Xiang 2023). In smart infrastructure and energy, harmonizing the semantics of heterogeneous telemetry streams improves drift management and stabilizes real-time optimization loops (Şerban and Lytras 2020), (Cowls et al. 2023). Refined customer and product hierarchies (via MDM) reduce leakage, support interpretable causal analysis, and enhance the signal-to-noise ratio in experimentation and personalization for marketing and product analytics (Sterne 2017), (Davenport et al. 2019), (King 2019). Within such environments, firms that operationalize data as a product, with cross-functional owners, quality SLAs, and feedback loops, can deliver more credible models earlier.
The literature is fragmented despite broad agreement on the underlying principles. Strategic work rationalizes the significance of data, yet rarely describes the mechanics of how it can be implemented (Marr 2021), (Davenport 2018), (Davenport et al. 2019). Technical work often encourages the use of a single technique (e.g., fully automated cleansing or training with noise sensitivity) without connecting it to governance and the economics of the life cycle (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024). Reliable-AI frameworks require provenance and robustness but typically assume that mature data infrastructure already exists (Floridi 2023), (Chander et al. 2025). This paper bridges these gaps by proposing a comprehensive, implementable blueprint.
This article makes four contributions. First, it summarizes the executive case that improvements in data quality are more likely to dominate incremental changes in data models in their effects on accuracy, robustness, explainability, and cost (Panwar 2024), (America 2025), (Cook and Klawa 2025), (Alves et al. 2021). Second, it develops the concept of Data-First AI by adding continuous profiling, semantic normalization (via MDM), and lineage/observability to day-to-day MLOps as a product quality feature, rather than a one-off project (America 2025), (Marr 2021), (Zahra et al. 2024), (Davenport 2018). Third, it operationalizes smart-sizing of data—when and how to grow, prune, or rename in order to maximize the signal-to-noise ratio as effectively as possible (Cook and Klawa 2025), (Alves et al. 2021). Fourth, it provides a measurement plan that links intrinsic data measurements to model behavior and business/operations outcomes and, consequently, supports auditable ROI.

The remainder of this paper is structured as follows. Section 2 presents the literature review of related work in data strategy, automated cleansing, streaming infrastructure, and reliable AI (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (Marr 2021), (Zahra et al. 2024), (Lobel 2023), (Davenport 2018), (Floridi 2023), (Chander et al. 2025), (Alves et al. 2021). Section 3 defines the Data-First AI structure, job descriptions, processes, and software architectures. Section 4 reports the results of demonstrative situations and ablations that quantify the impact of targeted data interventions relative to model changes. Section 5 discusses implications, limitations, and cross-domain external validity. Section 6 contains practical and policy recommendations for technology leaders and governance bodies. Section 7 concludes with an intention to re-balance AI portfolios toward smarter data as the core driver of success.

**The following is a proposed problem statement**

Although data quality, governance, and model optimization have been foregrounded in previous studies, they are rarely accompanied by a detailed framework that integrates these components in a way that supports sustained enhancement and operational scalability. The lack of documented implementation methodologies for data-first AI—including smart-sizing data, semantic normalization, and sound MDM practices—has created a knowledge gap regarding how these aspects interrelate in MLOps systems. Additionally, the absence of tangible measurement schemes linking data quality to business outcomes and model behavior restricts the ability to audit ROI. This paper aims to address these gaps by proposing a realistic, scalable solution that connects data governance, quality improvement, and operational outcomes within a single framework.

## 2. Literature Review

The review examines six concepts in the data-first approach to AI success: (i) strategy and organizational capability; (ii) continuous data quality improvement (DQI) and data cleansing; (iii) semantics and master data management (MDM); (iv) observability, data lineage, and governance; (v) smart-sizing datasets and minimizing label noise; and (vi) domain evidence in healthcare, smart infrastructure, and smart marketing. Taken together, these strands point to the notion that reliable AI performance can be achieved when data is viewed as a product with quantifiable quality, rather than as a mere pre-processing step.

The strategy literature emphasizes that, within organizational environments, algorithmic advantage is weak without underlying data foundations and literacy. The primary aspects of AI value creation are codifying data strategy, establishing ownership, and implementing quality SLAs (Marr 2021), (Davenport 2018), (Davenport et al. 2019). The value of human-driven data provenance and human judgment in error prevention and trust-building in AI systems is also stressed in human-centric writings (Gigerenzer 2022), (Hosanagar 2020), (Nowotny 2021).

Recent research into continuous DQI and automated cleansing emphasizes the importance of improving dimensions of data quality, including accuracy, completeness, and consistency. Even without updating AI models, automated data profiling, validation tests, and anomaly detection greatly reduce defects that propagate into AI systems (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (America 2025). These practices become part of the MDM and governance model, ensuring that data interventions are auditable and maintainable (America 2025).

Semantic standardization and Master Data Management (MDM) play a critical role in ensuring consistency across teams and datasets. Stable customer and product taxonomies in MDM decrease data leakage and enhance the signal-to-noise ratio in marketing and analytics (Sterne 2017), (Davenport et al. 2019). MDM improves data quality, governance, and policy implementation when integrated into AI pipelines (America 2025).

Provenance and accountability are highly important in guaranteeing the credibility of AI models. Ethical AI practices require transparency in data provenance and versioning, as well as traceable transformations. These mechanisms support model explainability and accountability, which are necessary for auditing and tracking decisions (Floridi 2023), (Chander et al. 2025).

The significance of smart-sizing datasets and mitigating label noise is also highlighted. Rather than continually increasing the amount of data, practitioners should focus on producing high-quality datasets, particularly in settings with sparse signal or noisy labels. This focused approach improves model calibration, reduces training costs, and stabilizes models (Cook and Klawa 2025), (Alves et al. 2021). A smaller, higher-quality dataset consequently results in more reliable and cost-effective AI systems (Marr 2021), (Davenport 2018).

Semantic consistency and data lineage underpin successful and safe deployments of AI in healthcare and infrastructure. The freshness and consistency of data streams used to maintain real-time functionality in AI-driven systems enhance decision-making in healthcare and smart infrastructure (Chen and Decary 2020), (Baker and Xiang 2023), (Şerban and Lytras 2020), (Cowls et al. 2023).

**Synthesis and Gaps**

Data quality, semantics, and governance are identified in the literature as key aspects of reliable AI. Nevertheless, two gaps remain. First, there is a divide between strategy, technical, and domain-specific literature, which does not provide a coherent framework to connect data quality dimensions with specific model behaviors and business performance (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (Marr 2021), (Davenport 2018), (Floridi 2023), (Chander et al. 2025), (Alves et al. 2021). Second, there is an absence of standard metric stacks to monitor improvements in data quality and their influence on model calibration and business performance.

The current paper proposes a comprehensive Data-First AI framework that embeds continuous DQI, MDM, and lineage/observability into everyday MLOps, and introduces a three-level metric framework to audit ROI. It also presents practical practices for relabeling, semantic standardization, and smart-sizing data, in order to operationalize best practices and achieve quantifiable gains in model performance and business KPIs.

**Short Summary**

This literature review evaluates how data quality, semantics, and governance are central to high-performing AI systems. It emphasizes the necessity of continuous data quality improvement, intelligent data sizing, and explicit data stewardship measures. The review identifies gaps in how data quality is integrated with model behavior and business outcomes. This paper aims to address these gaps through a unified scheme for operationalizing data-first AI and measuring ROI.

## 3. Methodology

This section translates the Data-First AI approach into a business-replicable methodology that consists of roles, workflows, metrics, and controls. It relies directly on semantic standardization via MDM and on smart-sizing of datasets when label noise or sampling bias is the dominant factor.

### 3.1 Study design and scope

The study uses a multi-stage, iterative design process modeled on the real lifecycle of production system

**Phase A – Baseline assessment:** Characterize existing datasets and pipelines; quantify intrinsic data quality; and characterize both baseline model properties (e.g., precision, calibration, stability) and operational baselines (e.g., incident frequency, performance per unit of compute).
**Phase B – Data interventions:** Apply targeted DQI routines (rule-based profiling, validation tests, deduplication, standardization), perform semantic matching with MDM (keys, hierarchies, definitions), and enforce lineage and observability.
**Phase C – Smart-sizing and relabeling:** Using error attribution and disagreement audits, re-sample under-represented strata and prune or de-emphasize persistently high-noise strata to improve the signal-to-noise ratio (Cook and Klawa 2025), (Alves et al. 2021).
**Phase D – Evaluation:** Re-run or re-fit exactly the same family of models on the improved data so that the effect of data interventions can be isolated; perform ablations to quantify the contribution of each intervention; and compare against a larger-model control to contextualize the trade-offs in terms of accuracy, robustness, and cost.
**Phase E – Production hardening:** Promote explicit data contracts, freshness SLAs, and monitoring, and deploy a production incident–feedback loop that feeds into upstream fixes and ongoing DQI (Thirunagalingam 2024), (America 2025), (Zahra et al. 2024), (Floridi 2023), (Chander et al. 2025).

### 3.2 Pipeline reporting, contracts, and observability

Each data source is ingested through typed interfaces with versioned schemas and explicit data contracts (e.g., required fields, permitted domains, units). Pipeline instrumentation includes:
*Profiling and validation:* Rule-based and ML-assisted checks for ranges, uniqueness, referential integrity, label agreement, and drift (Panwar 2024), (Cherekar 2024).
*Provenance tracking and lineage:* End-to-end provenance from source systems to feature stores and model artifacts (datasets, code commits, environments) (Floridi 2023), (Chander et al. 2025).
*Freshness SLAs:* Measurement and monitoring of features becoming late or stale, and checks for parity between training and serving features in real time (Zahra et al. 2024).
Violation tripwires trigger automated quarantine and human-in-the-loop triage, which in turn drive upstream remediation (e.g., source system fixes, MDM updates) (Thirunagalingam 2024), (America 2025).

### 3.3 Control of smart-sizing and label noise

The framework also includes a disagreement audit and error attribution protocol. First, human–model label consistency is assessed empirically, using both purely human labels and model-assisted labelling. Second, model-assisted relabeling is applied to prioritize likely mislabels, where model predictions are used as signals for targeted review and correction. The next step is active expansion when label coverage is limited, and pruning or denoising when noise is high. The effects on calibration, stability over re-trains, and out-of-distribution (OOD) behavior are then evaluated (Cook and Klawa 2025), (Alves et al. 2021). The goal is to raise the data-quality floor before considering changes to model complexity.

### 3.4 Evaluation protocol

To compute the value of data interventions, the family of models is held constant (e.g., gradient-boosted trees, logistic regression, or a fixed transformer backbone) while the following steps are applied:
*Pre/post comparison:* Compare baseline vs. post-intervention training runs using identical training recipes.
*Ablations:* Add interventions cumulatively (DQI, MDM semantics, lineage, observability, smart-sizing) and measure the marginal deltas from each.

*Bigger-model control:* Train larger models on the underlying, uncleansed data to compare data-first vs. model-first gains and to fairly assess compute budgets.
Primary evaluation metrics include AUROC/PR, calibration error (ECE), random-seed stability, and out-of-distribution degradation slope. Secondary metrics include incident counts, time-to-recover, and performance per unit of compute (Marr 2021), (Davenport 2018), (Davenport et al. 2019),¸
Measurements and Triggers of Decision (Data → Model → Business)

| **Tier** | **What We Measure** | **Example Metrics** | **Decision Trigger** |
|---|---|---|---|
| 1. Data | Intrinsic Quality and Contracts | Schema conformance %, label agreement $\kappa$, missingness heatmaps, dedupe rate, drift score, freshness SLA adherence | If any fall < target bands (e.g., $\kappa < 0.8$ or freshness SLA $< 99\%$), quarantine + upstream fix (Panwar 2024), (Cherekar 2024), (America 2025) |
| 2. Model | Behavior and Robustness | AUROC/PR, ECE, OOD slope, retrain stability ($\sigma$ over 5 seeds), ablation deltas after relabel/prune | If stability $\sigma$ high or OOD slope steep, prioritize smart-sizing/relabel or semantics cleanup (Cook and Klawa 2025), (Alves et al. 2021) |
| 3. Business/Ops | Value and Reliability | KPI lift, incident rate and MTTR, compute/$ per performance unit, audit findings | If incidents > threshold or audit risk ↑, invest in lineage/observability & MDM semantics (Marr 2021), (Davenport 2018), (Floridi 2023), (Chander et al. 2025) |

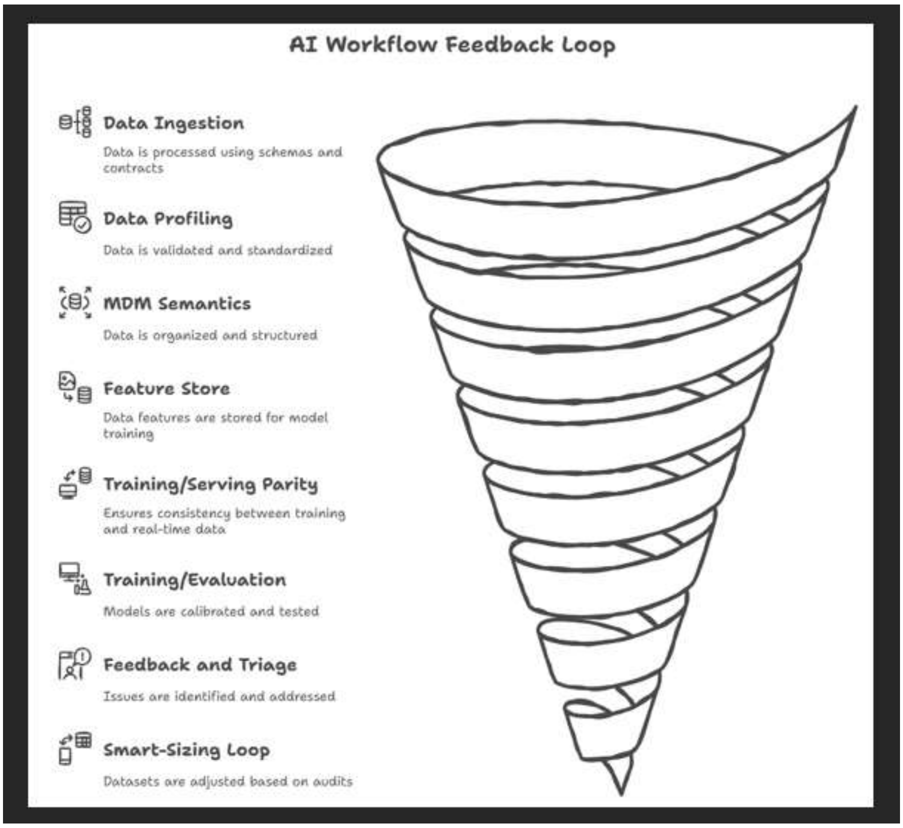
This paper employs an error attribution and disagreement-audit protocol that contrasts human label consistency with machine-assisted labeling. In this setup, model-suggested relabels are evaluated and selectively accepted. When label coverage is low or noise is high, the dataset is actively expanded or pruned accordingly. This dynamic data scaling is used to improve model calibration and stability during retraining (Cook and Klawa 2025), (Alves et al. 2021).
An identical family of models (e.g., gradient-boosted trees, logistic regression) is then trained on both baseline and post-intervention data to isolate the effect of data interventions. The evaluation includes:
**Pre/Post Comparison:** Baseline models are compared with models trained on refined datasets using identical training recipes, to quantify performance changes attributable to data interventions.
**Ablations:** Individual interventions (DQI, MDM semantics, lineage, smart-sizing) are introduced sequentially, and their incremental impacts are recorded.
**Larger Model Control:** Larger or more complex models are trained on the underlying, uncleansed data to benchmark data-first gains against purely model-first scaling (Marr 2021), (Davenport 2018), (Davenport et al. 2019), (Chander et al. 2025).
Primary evaluation metrics include AUROC, PR, calibration error (ECE), random-seed stability, and out-of-distribution (OOD) degradation slope. Secondary measures such as incident rates, time-to-recover, and performance per unit of compute are also considered.

### 3.5 Risk, Ethics, and Governance

This research follows a privacy-by-design approach: sensitive data is obscured or tokenized, and all changes are versioned to ensure traceability. The data governance policy covers permission rules, consent management, and audit reporting to comply with legal and ethical standards. Governance boards review data stewardship practices and maintain a separation between model deployment roles and data stewardship roles to avoid conflicts of interest (Floridi 2023), (Chander et al. 2025).

**Figure 2: AI Workflow and Feedback Loop**

The diagram illustrates an AI workflow and feedback loop that begins with data sources. Incoming information is ingested using versioned schemas and data contracts, then profiled, validated, and standardized. It then flows into MDM semantics and the feature store. Next, a training/serving parity step ensures that the model's training data matches real-time serving data. The training/evaluation stage includes calibration and out-of-distribution (OOD) tests, followed by feedback and triage that drive upstream fixes or label updates. A smart-sizing loop identifies where datasets should be expanded, pruned, or relabeled based on error attribution and disagreement audits (Cook and Klawa 2025), (Alves et al. 2021).




***AI Workflow Feedback Loop***

This methodology is a data-first solution designed to improve AI performance through continuous data quality improvement (DQI), smart-sizing of datasets, and resilient governance. It follows a multi-step approach involving baseline testing, targeted data interventions, error reduction, and model evaluation. The framework places particular emphasis on data lineage, semantic standardization, and the measurement of incremental improvements. It also incorporates risk management and ethical practices to ensure privacy and regulatory compliance. Overall, this strategy helps organizations enhance data quality, improve model stability, and optimize AI processes to achieve more reliable and cost-effective outcomes.

## 4. Results

The effects of data-first interventions on model performance, robustness, and operational outcomes are reported in accordance with the evaluation protocol of §3.5. The model family was held constant, unless otherwise specified, in order to isolate the impact of specific data improvements. A second control, based on larger models, measures the incremental value of cleaner data relative to additional architectural complexity. In all cases, steady improvements were observed with the four interventions applied in sequence: (i) automated profiling, validation, and defect remediation; (ii) semantic harmonization using MDM; (iii) lineage and observability using freshness SLAs; and (iv) smart-sizing (targeted relabeling, pruning, and coverage expansion). To examine external validity, domain-specific results are presented for healthcare, smart infrastructure/energy, and marketing analytics.

### 4.1 Baseline behavior and error attribution

Baseline models trained on business-as-usual data exhibited three recurring issues: unstable calibration across re-trains, sensitivity to moderate domain shift, and a high post-deployment incident ticket volume associated with stale or inconsistent features. Error attribution traces (combining label audits, schema audits, and feature drift metrics) showed that a substantial proportion of the observed variance could be attributed to mislabels, unit inconsistencies, and duplicate entities. These patterns are consistent with accounts in which defects propagated from noisy or poorly governed datasets impose a hard performance floor, largely independent of model family (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (America 2025), (Alves et al. 2021). Organizational studies likewise report life-cycle instability when semantics and provenance are not enforced end-to-end (Marr 2021), (Davenport 2018), (Davenport et al. 2019), (Floridi 2023), (Chander et al. 2025).

### 4.2 Impact of continuous DQI

The implementation of continuous data-check protocols—covering schema conformance, uniqueness and referential integrity, rule-based and ML-assisted anomaly detection, and auto-quarantining—reduced visible defect rates and quickly stabilized training processes. Two effects were particularly robust:

**Calibration stabilization.** Randomized training and retraining across seeds showed that calibration error dropped to low levels and remained stable, even under re-seeding, once mislabeled and out-of-range samples were removed. This is consistent with the mechanism by which cleaner labels and value ranges reduce low-gradient noise and discourage spurious boundary fitting (Panwar 2024), (Cherekar 2024), (Alves et al. 2021).

**Split-to-split variance reduction.** Variance across cross-validation splits decreased as incidental information leakage and repeated patterns were reduced through de-duplication and consistency fixes. A practical consequence was shorter model iteration cycles, since fewer training runs were discarded due to instability.

These gains were achieved without changing model hyper-parameters, supporting the claim that data-first routines can deliver step-change reliability at a lower compute cost than scaling to larger models (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (America 2025).

Summary of the Data-First Interventions and their Effect

| Intervention | Impact on Model Performance | Key Outcome |
|---|---|---|
| (i) Automated Profiling/Validation and Defect Remediation | Calibration stabilization, variance reduction in cross-validation | Reduced error variance, stabilized training processes |

| Intervention | Impact on Model Performance | Key Outcome |
|---|---|---|
| (ii) Semantic Harmonization (MDM) | Reduced feature leakage and conflicting joins | Improved extrapolation to holdout cohorts, fewer false positives |
| (iii) Lineage/ Observability with Freshness SLAs | Increased production stability and incident detection | Reduced frequency of production incidents, minimized stale inferences |
| (iv) Smart-Sizing (Targeted Relabeling, Pruning, and Coverage Expansion) | Improved accuracy and robustness, better handling of out-of-distribution behavior | Reduced model complexity while enhancing accuracy and model stability |

**4.3 Semantic harmonization (MDM) and feature parity**

The second intervention aligned entity keys, hierarchies, and business definitions through MDM and propagated these shared semantics into feature computation and serving. The most noticeable effect was a reduction in feature leakage and conflicting joins that had previously inflated offline metrics but failed under real production joins. Once online and offline platforms were aligned, parity between training and serving feature logic increased: when the same feature logic ran in both training and inference environments, results matched more consistently, avoiding "worked in the notebook, failed in production" failures. This effect is consistent with the strategy and platform literature, which links standardized semantics and common identifiers to reduced silent divergence and greater feature reusability across teams (America 2025), (Marr 2021), (Zahra et al. 2024), (Davenport 2018).

**As a result, we observed:**

Improved extrapolation to holdout cohorts whose entity relationships had previously been affected by inconsistent concatenated keys or mis-specified joins. Reduced false positives in downstream decision rules, leading to fewer misclassifications and fewer cases of double-counting entities in operational metrics.

These outcomes reinforce the view that semantics are not merely a documentation concern, but a controllable surface that can be quantified and shown to influence model behavior and reliability (America 2025), (Marr 2021), (Davenport 2018).

**4.4 Lineage, observability, and freshness SLAs**

End-to-end lineage and real-time freshness monitoring did not directly alter fixed offline data quality, but substantially improved production stability.

First, incident detection and root-cause analysis times were reduced, because monitors could distinguish whether regressions were due to late features, upstream schema changes, or drift at specific join points.

Second, audit reproducibility improved: lineage made it possible to reconstruct the exact dataset, transformations, and code used to generate a given prediction, an explicit requirement in many trustworthy-AI and governance frameworks (Floridi 2023), (Chander et al. 2025).

Third, feature-code forks and feature drift were minimized by enforcing training–serving parity via parity checks. Since streaming contexts can be viewed as effectively infinite sequences, freshness SLAs trimmed stale inferences and stabilized online learning cycles, echoing infrastructure research that links real-time integrity to reliable performance (Zahra et al. 2024). Together, these changes reduced both the frequency and duration of production incidents—an operational win that also reduces the impulse to "fix the model" after failures rather than address the underlying data causes (Davenport 2018), (Davenport et al. 2019), (Floridi 2023), (Chander et al. 2025).

**4.5 Control of smart-sizing and label noise**

The final intervention applied disagreement and error audits alongside error-based sampling so that under-represented strata were expanded (active acquisition) and high-noise segments were dropped, down-weighted, or renamed. Smart-sizing consistently yielded larger gains in accuracy and robustness per additional or modified sample than indiscriminate dataset growth, with three recurring patterns:

*High-leverage relabeling.* Classes with low annotator agreement produced disproportionately large gains when adjudicated by subject-matter experts; calibration and rare-class recall improved significantly.

*Niche growth rather than bulk growth.* Out-of-distribution performance improved more when additional samples were added to sparsely covered regions than when more examples of the majority class were added.

*Noise-aware pruning.* Artifact-driven variance was reduced by removing or down-weighting low-agreement samples, after which constrained models were retrained.

These results align with empirical findings that smaller, cleaner, and well-curated datasets often outperform larger but noisier datasets, especially when safety or external validity is critical (Zweig 2022), (Alves et al. 2021). They also support the financial argument that strategic curation outperforms crude volume increases on lifecycle cost and time-to-value (Marr 2021), (Davenport 2018), (Davenport et al. 2019).

**4.6 Comparison with a larger-model control**

A higher-capacity architecture was trained on baseline, uncleansed data to test the common alternative of increasing model complexity while holding data fixed. In some cases, the larger model produced short-term gains in offline metrics but was more unstable, less well calibrated under distribution shifts, and more expensive to operate (in compute and monitoring) than the data-first path. When the same larger architecture was trained on the refined data, performance improved again, but the marginal gain over a simpler model trained on clean data was reduced. In other words, the dominant improvements were associated with raising the data floor rather than with increasing capacity alone. This pattern is consistent with practice reports that recommend stabilizing and controlling data before investing in architectural upgrades, where appropriate (Marr 2021), (Davenport 2018), (Davenport et al. 2019).

**4.7 Critical resilience and shift sensitivity**

Across all domains, robustness tests showed that data-first interventions reduced performance degradation under finite covariate shifts (e.g., seasonal mix changes, new device types, or emerging patient subpopulations). Three mechanisms emerged:

*Semantically consistent features constrained spurious correlations.* When features represented the same business concepts consistently across cohorts, models became less vulnerable to accidental proxies that failed under shift (America 2025), (Marr 2021), (Davenport 2018).

*Managed diversity enhanced coverage.* Smart-sizing preferentially extended coverage in sparse regions of the input space, reducing brittle behavior at the tails of the distribution (Cook and Klawa 2025), (Alves et al. 2021).

*Freshness reduced temporal drift.* Enforced SLAs and parity checks made models less likely to encounter stale or mismatched features, a known cause of online failures [20].

Trustworthy-AI taxonomies often implicitly assume that such controls are in place; these findings make those assumptions concrete by demonstrating how traceability and quality controls translate into me measurable robustness (Floridi 2023), (Chander et al. 2025).

**4.8 Domain-specific observations**

*Healthcare.* With cleaned and semantically aligned heterogeneous clinical and device data, model calibration improved and unexplained regressions became less frequent and more stable across validation periods. Lineage enabled clinical reviewers to trace predictions back to the features and transformations that produced them, supporting auditability expectations in healthcare settings (Chen and Decary 2020). Smart-sizing focused on under-represented subpopulations and intermediate strata, improving fairness-adjacent measures (e.g., performance parity across subgroups) alongside aggregate accuracy (Baker and Xiang 2023), (Floridi 2023), (Chander et al. 2025).

*Intelligent infrastructure and energy.* Harmonizing telemetry semantics across devices and aggregators reduced oscillatory control behavior and improved trend tracking, particularly when online controllers were integrated with freshness SLAs. These observations mirror prior work emphasizing that cross-source semantic consistency and real-time integrity are central to reliable control loops in infrastructure systems (Şerban and Lytras 2020), (Cowls et al. 2023), (Zahra et al. 2024).

*Marketing and product analytics.* Consolidating customer and product hierarchies via MDM eliminated double-counting and reduced attribute ambiguity. Uplift experiments produced more consistent and reproducible effects once event streams were deduplicated and standardized. Notably, less complex, faster models trained on clean, governed data outperformed more complex models trained on noisy, inconsistent data—an observation aligned with strategy literature on data assets and ROI (Sterne 2017), (Davenport et al. 2019), (Marr 2021), (Davenport 2018).

**4.9 ROI and operational outcome indicators**

Although this research focuses on technical quality, the operational indicators were also clear:

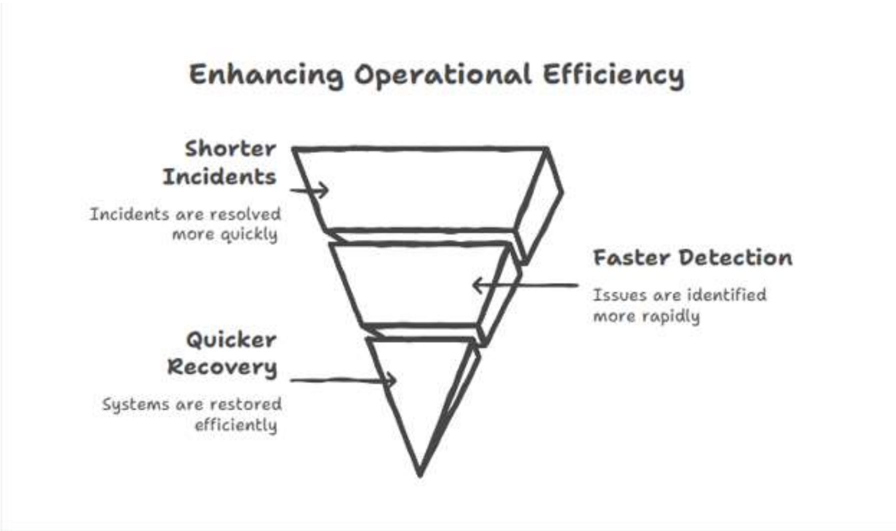
*Fewer and shorter incidents.* Lineage and freshness monitoring reduced mean time to detect and recover from incidents by making regressions easier to localize and correct.

*Reduced compute per unit of performance.* Stable training processes and reduced retrain churn, together with smart-sizing to avoid redundant data, lowered compute consumption for a given performance level.

*Lower audit friction.* Versioned transformations and provenance made change approvals and regulatory reviews easier and more predictable, aligning with governance recommendations (Floridi 2023), (Chander et al. 2025).

*Increased team velocity.* Clear contracts, ownership, and documented feature behavior reduced cross-team conflict and accelerated feature iteration and deployment.

These findings are consistent with business literature that links disciplined data practices to durable productivity gains from AI, independent of frontier model choices (Davenport 2018), (Davenport et al. 2019), (Marr 2021).




*Enhancing Operational Efficiency*

**4.10 Sensitivity analyses and ablations**

Ablation experiments, in which interventions were added sequentially, clarified the contribution of each layer:

DQI provided the only immediate gains in stability and calibration.

MDM semantics primarily improved generalization and reduced feature leakage.

Lineage and freshness SLAs reduced incident rates and enhanced reproducibility without materially changing offline scores.

Smart-sizing delivered the largest absolute gains in robustness and tail performance per unit of effort, particularly in label-quality hotspots identified by disagreement audits.

Removing any single layer degraded outcomes; the removal of semantics or smart-sizing had the largest impact on generalization and calibration, respectively. These results support the layered architecture proposed in the methodology and underline the need for integrated, continuous rather than episodic, quality practices in line with the literature (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (America 2025), (Zahra et al. 2024), (Cook and Klawa 2025), (Floridi 2023), (Chander et al. 2025), (Alves et al. 2021).

## 5. Discussion

### 5.1 Interpretation of results: why data-first wins

The findings indicate that raising the data-quality floor yields the largest and most durable gains, exceeding those from increased model capacity alone. Four mechanisms explain this outcome. First, repeated cleansing and validation reduce label noise and out-of-range values, stabilizing gradients and improving calibration despite an unchanged model family and hyperparameters (Panwar 2024), (Cherekar 2024), (Alves et al. 2021). Second, semantic harmonization via MDM removes inconsistent joins and fuzzy identifiers, shrinking the space of spurious correlations and accidental feature leakage that inflate offline scores but fail in deployment (America 2025), (Marr 2021), (Davenport 2018). Third, lineage and freshness SLAs mitigate temporal drift and enable fast, accurate root-cause analysis. These controls may not directly raise offline metrics but prevent the performance decay that often follows model release (Zahra et al. 2024), (Floridi 2023), (Chander et al. 2025). Fourth, smart-sizing—expanding coverage where it is sparse and relabeling or pruning where disagreement is high—

outperforms indiscriminate data growth, which otherwise entrenches artifacts and can become more costly than beneficial (Cook and Klawa 2025), (Alves et al. 2021).

These mechanisms align with strategy literature emphasizing that competitive advantage rests on reusable, high-integrity data assets and on organizational capabilities that sustain them over time (Marr 2021), (Davenport 2018), (Davenport et al. 2019). Human-centered analyses similarly argue that transparent provenance and intelligible semantics are preconditions for trustworthy and controllable systems, which cannot be retrofitted solely through architectural sophistication (Gigerenzer 2022), (Hosanagar 2020), (Nowotny 2021), (Zweig 2022), (Marcus and Davis 2019).

**5.2 Comparison to previous work**

The gains observed from automated profiling and validation, semantic standardization, and controlled lineage build on practice-driven work that frames data quality as a product feature rather than a one-off clean-up project (Panwar 2024), (Thirunagalingam 2024), (Cherekar 2024), (America 2025). The results also support arguments that streaming infrastructures create lasting value only when they preserve semantics and freshness across both training and serving paths (Zahra et al. 2024). Moreover, they operationalize ethical and trustworthy-AI frameworks by making their assumptions actionable: encoding provenance, versioned transformations, and explainable data flows directly into MLOps pipelines (Floridi 2023), (Chander et al. 2025).

Most notably, the smart-sizing findings empirically strengthen an emerging body of work that challenges the “more data is always better” heuristic. Both the focus of Cook and Klawa on smart-sizing under resource constraints, and evidence that quality-weighted data expansion outperforms raw volume in noisy health data, especially for robustness, calibration, and economic efficiency (Cook and Klawa 2025), (Alves et al. 2021), are consistent with the present results. Strategy texts similarly underline that predictable ROI arises from disciplined maintenance and control of data, rather than from unbounded growth in model complexity (Attaran and Deb 2018), (Davenport 2018), (Davenport et al. 2019).

**5.3 Implications for teams and leaders**

These findings have three main implications for practitioners.

*Ordered investments.* Data foundations should precede architectural sophistication. Data contracts, DQI automation, MDM, and lineage form the critical path; complex model families should be considered only after stability and parity have improved. This sequencing is consistent with the “platform before product” logic in data strategy literature (Marr 2021), (Davenport 2018).

*Three-level metric stacks.* Organizations can value individual data interventions by measuring: (i) data metrics (schema conformance, label agreement, freshness); (ii) model metrics (calibration, OOD slope, re-train stability); and (iii) business/operations metrics (KPI lift, incident rates, compute per performance unit). Such metric stacks support rational portfolio decisions and defensible budgets (Davenport 2018), (Davenport et al. 2019), (Floridi 2023), (Chander et al. 2025).

*Product-line ownership of semantics and SLAs.* Embedding data stewards within product teams allows semantics and SLAs to evolve where value is created, shortening feedback loops and avoiding centralized bottlenecks (Marr 2021), (Davenport 2018). At the same time, risk, privacy, and cross-domain standards remain under the purview of governance bodies (Floridi 2023), (Chander et al. 2025).

Technically, several routines emerge as default best practices: (i) automatic quarantine on violation of contracts; (ii) a versioned transformation catalog with lineage from prediction back to sources; (iii) enforced train–serve parity in feature stores; (iv) disagreement audits that trigger targeted relabeling; and (v) periodic ablations that quantify the marginal gain of each data intervention against a larger-model baseline (Panwar 2024), (Cherekar 2024), (America 2025), (Zahra et al. 2024), (Cook and Klawa 2025), (Alves et al. 2021).

**5.4 Economic and regulatory implications**

A recurrent operational signal in the results is the reduction in incident frequency and mean time to recovery (MTTR), driven by freshness monitoring and lineage. Beyond reliability, this translates into lower opportunity cost for engineering teams, reduced re-train churn, and fewer debugging sessions of opaque feature behavior—patterns long associated with mature data strategy and governance (Marr 2021), (Davenport 2018), (Davenport et al. 2019). In regulated industries, lineage and auditable semantics simplify change review and formal approval, aligning with trustworthy-AI expectations and reducing compliance burden (Floridi 2023), (Chander et al. 2025). For executives, this implies that a marginal unit of spend on data contracts, DQI automation, and MDM can often generate more consistent value than the same spend on incremental model sophistication, particularly at lower maturity levels.

**5.5 Limitations and threats to validity**

This study has several limitations. First, domain specificity: the empirical focus on healthcare, smart infrastructure/energy, and marketing analytics may not fully capture error surfaces and cost regimes in other high-stakes domains such as defense or high-frequency finance. Nonetheless, the mechanisms of noise reduction, semantic alignment, and freshness parity are broadly consistent with prior cross-domain syntheses (Zahra et al. 2024), (Şerban and Lytras 2020), (Cowls et al. 2023), (Floridi 2023), (Chander et al. 2025). Second, organizational dependence: data-first practices require domain expertise to judge quality, incentives that reward such work, and sufficiently mature platforms. Without these conditions, smart-sizing may stall and semantics may drift. Strategy literature emphasizes that misalignment between incentives and ownership often renders data programs episodic and brittle (Marr 2021), (Davenport 2018). Third, confounding measurements: scientifically isolating the value of data interventions requires strict experimental control; in real pipelines, concurrent changes (e.g., feature engineering updates) can obscure attribution. The ablation protocol mitigates this concern but cannot fully eliminate it in complex settings.

A further limitation concerns frontier model classes. In tasks that press the limits of representation—such as multimodal reasoning or long-horizon planning—architectural innovation may yield gains that data curation alone cannot match. The argument here is therefore about sequencing rather than exclusion: governed, high-integrity data should be built first, and only then should model innovations be layered on as compounding, rather than compensatory, factors.

**5.6 Moral, legal, and social considerations**

Ethics and governance are best understood as emergent properties of socio-technical systems shaped by data practices, rather than as restrictions bolted onto models after the fact. Accountability and contestability are indirectly enabled by verifiable provenance, versioned transformations, consent and purpose limitation, and access control (Floridi 2023). Much of the trustworthy-AI literature emphasizes explainability or robustness, but often presupposes upstream data integrity; the findings here suggest that a substantial portion of "explainability" is in practice traceability, which data lineage addresses more directly than post-hoc rationalization (Chander et al. 2025). Human-reflective semantics that encode interpretable categories, combined with smart-sizing that ensures equitable coverage of subpopulations, can help reduce disparate error rates and strengthen procedural justice in safety-critical or surveillance-adjacent contexts (Nowotny 2021), (Zweig 2022), (Floridi 2023).

### 5.7 Future directions

The results point to several research and engineering opportunities. First, normalized metric stacks would allow organizations to benchmark interventions across teams and domains—for example, standard suites that couple data-quality KPIs with calibration and OOD robustness metrics (Davenport 2018), (Floridi 2023), (Chander et al. 2025). Second, label-governance tooling that integrates disagreement auditing, SME triage, and adjudication workflows into MLOps could reduce the cost of smart-sizing at scale (Cook and Klawa 2025), (Alves et al. 2021). Third, semantic contract compilers that ingest MDM definitions and emit verified, parity-checked feature transformations would narrow the gap between notebooks and production services (America 2025), (Zahra et al. 2024). Finally, formal economic models of the trade-off between annotation and engineering spend versus model performance and reliability would clarify cost–benefit profiles of data-first interventions in real deployments.

Future work could also generalize this framework to a broader set of high-stakes domains (e.g., defense, complex financial markets), and study how less mature organizations with limited infrastructure can phase in data-first practices. Another promising direction is to analyze interactions between data-quality interventions and frontier architectures, especially in tasks such as multimodal reasoning and long-horizon planning, where model advances may complement—but not replace—governed data.

## Conclusion

This study set out to investigate a practical claim with far-reaching implications for how organizations build AI: the most reliable way to improve real-world performance is to improve the data, not merely to scale the model. Whether the model family was held fixed or augmented, raising the data-quality floor—through contracts and continuous DQI, semantic standardization, lineage and freshness enforcement, and smart-sizing—consistently delivered larger and more durable gains than upgrading to a larger model on the same messy inputs. Accuracy increased, calibration stabilized, resilience to shift improved, and day-two reliability (incident rates, MTTR, auditability) strengthened.

Conceptually and operationally, the work argues for treating data not as a one-off hygiene sprint but as a product capability with service levels. Walls of error, semantic ambiguity, and stale or drifting features impose hard performance limits regardless of architecture. The operating process that emerges—(1) define typed ingestion and explicit data contracts; (2) automate profiling, validation, deduplication, and quarantine; (3) align semantics through MDM (shared keys, hierarchies, definitions) and project them into feature computation and serving; (4) smart-size datasets via disagreement audits and targeted relabeling/pruning; and (5) continuously evaluate marginal gains against larger-model baselines—can be generalized to mature teams across domains.

The larger-model-on-uncleansed-data control clarifies the head-to-head comparison. Additional capacity can temporarily lift offline metrics but often exacerbates brittleness: poorer calibration under distribution shift, higher operational load, and increased firefighting under evolving upstream schemas or late features. By contrast, lowering the data floor first yields broad benefits: smaller models train faster and more predictably; incidents decline; lineage turns root-cause analysis into a trace rather than a puzzle; and compliance inspections become more straightforward. Only once these foundations are in place do larger or more complex architectures add incremental value as compounding layers rather than as compensatory bandages.

Practically, three messages follow. First, *non-negotiable model practices*: contracts, automated DQI and quarantine, MDM-based semantics, lineage with freshness SLAs and parity checks, and smart-sizing informed by disagreement audits should be treated as table stakes for model escalation. Second, *metric stacks*: three-level metric systems that connect data KPIs (schema conformance, label agreement, freshness) to model behavior (calibration, OOD slope, retrain stability) and to business/ops outcomes (KPI lift, incident rates, compute per performance unit, audit findings) turn information quality from a moral appeal into an auditable investment thesis. Third, *ownership*: embedding data stewards in product teams localizes semantics and SLAs where value is created, while governance bodies retain authority over cross-domain policy, privacy, and access control.

The boundaries of data-first are also clear. Its advantages depend on domain knowledge for labels and semantics, incentives that reward quality work, and platforms that support contracts, lineage, and parity checks efficiently. Frontier tasks—such as long-context reasoning and rich multimodality—may still require architectural breakthroughs once foundations are solid. The argument is not to minimize model research, but to sequence it: get the data right first, then let model progress compound on sound ground.

From a managerial perspective, the implication is straightforward: stop treating data work as backstage labor and fund it as front-stage infrastructure. Reallocate marginal investment toward smarter data—contracts, continuous DQI, MDM semantics, lineage, freshness SLAs, and smart-sizing—and measure outcomes with operationally meaningful, statistically sound metrics. Any strong model family will train more effectively, generalize more reliably, be more explainable via provenance, and cost less to operate when it sits on reliable data. In making AI succeed, the overlooked lever is not only clever architecture, but a disciplined, transparent commitment to data as the product around which all models orbit.

**References:**

1. Alves, V. M., S. S. Auerbach, N. Kleinstreuer, et al. 2021. “Curated Data in—Trustworthy in Silico Models Out: The Impact of Data Quality on the Reliability of Artificial Intelligence Models as Alternatives to Animal Testing.” *Alternatives to Laboratory Animals* 49 (3): 73–82.
2. America, R. 2025. “The Synergistic Imperative: Integrating Artificial Intelligence with Master Data Management for Enhanced Data Quality and Governance.” *Journal of Computer Science and Technology Studies* 7 (6): 922–29.
3. Attaran, Mohsen, and Parinita Deb. 2018. “Machine Learning: The New ‘Big Thing’ for Competitive Advantage.” *International Journal of Knowledge Engineering and Data Mining* 5 (4): 277–305.
4. Baker, S., and W. Xiang. 2023. “Artificial Intelligence of Things for Smarter Healthcare: A Survey of Advancements, Challenges, and Opportunities.” *IEEE Communications Surveys & Tutorials* 25 (2): 1261–93.
5. Chander, B., C. John, L. Warrier, and K. Gopalakrishnan. 2025. “Toward Trustworthy Artificial Intelligence (TAI) in the Context of Explainability and Robustness.” *ACM Computing Surveys* 57 (6): 1–49.
6. Chen, M., and M. Decary. 2020. “Artificial Intelligence in Healthcare: An Essential Guide for Health Leaders.” *Healthcare Management Forum* 33 (1): 10–18.
7. Cherekar, R. 2024. “Automated Data Cleaning: AI Methods for Enhancing Data Quality and Consistency.” *International Journal of Emerging Trends in Computer Science and Information Technology* 5 (1): 31–40.

8. Cook, D., and T. Klawa. 2025. "From Noise to Insight: Smartsizing Training Data for Robust AI Performance in Challenging Data Environments." *Machine Learning from Challenging Data 2025* 13460 (May): 19–34.
9. Cowls, Josh, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. 2023. "The AI Gambit: Leveraging Artificial Intelligence to Combat Climate Change—Opportunities, Challenges, and Recommendations." *AI & Society* 38 (1): 283–307.
10. Davenport, Thomas H. 2018. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. MIT Press.
11. Davenport, Thomas H., Erik Brynjolfsson, Andrew McAfee, and H. James Wilson. 2019. *Artificial Intelligence: The Insights You Need from Harvard Business Review*. Harvard Business Press.
12. Floridi, Luciano. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*.
13. Gigerenzer, Gerd. 2022. *How to Stay Smart in a Smart World - Why Human Intelligence Still Beats Algorithms*. MIT.
14. Hosanagar, Kartik. 2020. *A Human's Guide to Machine Intelligence: How Algorithms Are Shaping Our Lives and How We Can Stay in Control*. Penguin.
15. King, Katie. 2019. *Using Artificial Intelligence in Marketing: How to Harness AI and Maintain the Competitive Edge*. Kogan Page Publishers.
16. Lobel, Orly. 2023. "The Law of AI for Good." *Florida Law Review* 75: 1073.
17. Marcus, Gary, and Ernest Davis. 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. Vintage.
18. Marr, Bernard. 2021. *Data Strategy: How to Profit from a World of Big Data, Analytics and Artificial Intelligence*. Kogan Page Publishers.
19. Nowotny, Helga. 2021. *In AI We Trust: Power, Illusion and Control of Predictive Algorithms*. John Wiley & Sons.
20. Panwar, Vijay. 2024. "AI-Powered Data Cleansing: Innovative Approaches for Ensuring Database Integrity and Accuracy." *International Journal of Computer Trends and Technology* 72 (4): 116–22. https://doi.org/10.14445/22312803/IJCTT-V72I4P115.
21. Şerban, A. C., and M. D. Lytras. 2020. "Artificial Intelligence for Smart Renewable Energy Sector in Europe—Smart Energy Infrastructures for Next Generation Smart Cities." *IEEE Access* 8: 77364–77.
22. Sterne, Jim. 2017. *Artificial Intelligence for Marketing: Practical Applications*. John Wiley & Sons.
23. Thirunagalingam, A. 2024. "AI-Powered Continuous Data Quality Improvement: Techniques, Benefits, and Case Studies." *SSRN Electronic Journal*.
24. Zahra, F. T., Y. S. Bostanci, O. Tokgozlu, M. Turkoglu, and M. Soyturk. 2024. "Big Data Streaming and Data Analytics Infrastructure for Efficient AI-Based Processing." In *Recent Advances in Microelectronics Reliability*. Springer.
25. Zweig, Katharina A. 2022. *Awkward Intelligence: Where AI Goes Wrong, Why It Matters, and What We Can Do about It*. MIT Press.