

Review

Data Quality in the Age of AI: A Review of Governance, Ethics, and the FAIR Principles

Miriam Guillen-Aguinaga ¹, Enrique Aguinaga-Ontoso ^{2,3} , Laura Guillen-Aguinaga ⁴,
Francisco Guillen-Grima ^{5,6,7,8,*}  and Ines Aguinaga-Ontoso ^{6,7,8,*} 

¹ School of Law, International University of La Rioja, 26006 Logroño, Spain; miriamguillenaguinaga@gmail.com

² Department of Sociosanitary Sciences, University of Murcia, 30120 Murcia, Spain; aguinaga@um.es

³ Department of Preventive Medicine, Virgen de la Arrixaca University Clinical Hospital, 30120 Murcia, Spain

⁴ Department of Nursing, Clinica Universidad de Navarra, 28027 Madrid, Spain; lguillen@alumni.unav.es

⁵ Department of Preventive Medicine, Clinica Universidad de Navarra, 31008 Pamplona, Spain

⁶ Department of Health Sciences, Public University of Navarra, 31008 Pamplona, Spain

⁷ Group of Clinical Epidemiology, Area of Epidemiology and Public Health, Healthcare Research Institute of Navarre (IdiSNA), 31008 Pamplona, Spain

⁸ CIBER in Epidemiology and Public Health (CIBERESP), Institute of Health Carlos III, 46980 Madrid, Spain

* Correspondence: f.guillen.grima@unavarra.es (F.G.-G.); ines.aguinaga@unavarra.es (I.A.-O.)

Abstract

Data quality is fundamental to scientific integrity, reproducibility, and evidence-based decision-making. Nevertheless, many datasets lack transparency in their collection and curation, undermining trust and reusability across research domains. This narrative review synthesizes scientific and technical literature published between 1996 and 2025, complemented by international standards (ISO/IEC 25012, ISO 8000), to provide an integrated overview of data quality frameworks, governance, and ethical considerations in the era of Artificial Intelligence (AI). Sources were retrieved from PubMed, Scopus, Web of Science, and grey literature. Across sectors, accuracy, completeness, consistency, timeliness, and accessibility consistently emerged as universal quality dimensions. Evidence from healthcare, business, and public administration suggests that poor data quality leads to substantial financial losses, operational inefficiencies, and erosion of trust. Emerging frameworks are increasingly integrating FAIR principles (Findability, Accessibility, Interoperability, Reusability) and incorporating ethical safeguards, including bias mitigation in AI systems. Data quality is not solely a technical issue but a socio-organizational challenge that requires robust governance and continuous assurance throughout the data lifecycle. Embedding quality and ethical governance into data management practices is crucial for producing trustworthy, reusable, and reproducible data that supports sound science and informed decision-making.

Keywords: data quality; data governance; FAIR principles; ISO/IEC 25012; data lifecycle; artificial intelligence ethics; healthcare data



Academic Editor: Jamal
Jokar Arsanjani

Received: 17 September 2025

Revised: 11 November 2025

Accepted: 24 November 2025

Published: 4 December 2025

Citation: Guillen-Aguinaga, M.; Aguinaga-Ontoso, E.; Guillen-Aguinaga, L.; Guillen-Grima, F.; Aguinaga-Ontoso, I. Data Quality in the Age of AI: A Review of Governance, Ethics, and the FAIR Principles. *Data* **2025**, *10*, 201. <https://doi.org/10.3390/data10120201>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In today's data-driven world, data quality has become a decisive factor in advancing science, ensuring reliable decision-making, and achieving organizational success. In healthcare, business, and public administration alike, low-quality data can lead to flawed conclusions, wasted resources, and significant social and economic harm [1–4]. Data quality is commonly understood as the extent to which data are “fit for use” in supporting

operations, planning, and decision-making [5,6]. This definition highlights that quality is not an absolute property but a context-dependent attribute that must be assessed relative to specific goals and users.

Frameworks such as ISO/IEC 25012 and ISO 8000 define data quality in terms of dimensions, including accuracy, completeness, and consistency [7–9].

Most studies treat technical, ethical, and sectoral aspects of data quality separately. Integrated approaches remain scarce [10–12]. This fragmentation hinders practitioners who must navigate disparate studies to implement comprehensive data management practices.

This review synthesizes literature on data quality across three domains: standards, sectoral applications (e.g., healthcare, CRM), and emerging links with governance, ethics, and AI.

This principle is critically important for research, as ensuring data quality is a prerequisite for transparency, reproducibility, and accountability in the scientific process. Robust data quality practices enhance the traceability of datasets and the credibility of scientific conclusions, aligning with broader initiatives such as the FAIR principles, which stipulate that data should be Findable, Accessible, Interoperable, and Reusable.

This narrative review offers a unique contribution by synthesizing three areas that are seldom connected: established theoretical frameworks, practical applications in the business world, and emerging trends such as ethical governance and the use of Artificial Intelligence (AI) in data management. Unlike works that focus on a single dimension, this review offers a holistic perspective that connects technical challenges with their organizational and ethical implications, underscoring their direct impact on scientific reproducibility and strategic decision-making.

Data quality is typically assessed through a set of measurable dimensions—accuracy, completeness, consistency, timeliness, relevance, and validity—which collectively provide a framework for identifying and addressing potential issues [10,13,14]. These dimensions are widely applied across domains, including Customer Relationship Management (CRM). In SMEs, for example, failures in CRM implementation have often been linked to poor data quality—particularly in terms of accuracy, consistency, and completeness—underscoring the importance of robust evaluation tools and processes [12].

2. Material and Methods

2.1. Review Design and Reporting

We conducted a comprehensive narrative literature review to synthesize the evolution of data quality concepts, frameworks, and practices, explicitly bridging academic theory and industry- and standards-based practices. Reporting adheres to good-practice elements for transparent search descriptions (PRISMA-S items, where applicable). The time window was 1996–2025, chosen to capture foundational work (e.g., the consumer-centric “fitness for use” perspective emerging in the mid-1990s) and contemporary developments in AI, ethics, and standards; searches were last updated on 8 October 2025. No protocol was registered, given the narrative (non-systematic) design.

Rationale for Methodological Choices

We adopted a narrative rather than a systematic review design for several reasons. First, the field of data quality spans diverse domains (healthcare, business, public administration, software engineering) with heterogeneous methodologies that preclude meta-analytic synthesis. Second, our objective was conceptual integration—connecting theoretical frameworks (ISO standards, academic models) with practical applications and emerging trends (AI, ethics, FAIR)—rather than estimating quantitative effects. Third, we deliberately included grey literature (ISO/IEC standards, technical reports) alongside peer-reviewed

articles to capture industry practices and normative frameworks that systematic reviews typically exclude.

The 1996–2025 timeframe was chosen strategically to encompass three critical periods: (1) the foundational consumer-centric perspective of data quality [15]; (2) the formalization of international standards (ISO 8000 series, 2011–2015; ISO/IEC 25012); and (3) contemporary developments in AI ethics, governance frameworks, and FAIR principles for research data (2015–2025). This 29-year window captures both the field’s historical evolution and its current state.

Database selection balanced comprehensiveness with relevance. PubMed/MEDLINE provides coverage of biomedical and health sciences, while Scopus offers a multidisciplinary breadth with strong indexing of computer science and business literature. Web of Science ensures access to high-impact journals across disciplines, and Google Scholar facilitates the discovery of grey literature, preprints, and non-indexed sources. Together, these sources cover the theoretical, applied, and standards-based literature essential to our synthesis.

2.2. Information Sources

We searched four sources: PubMed/MEDLINE, Scopus, Web of Science Core Collection (WoS), and Google Scholar. To ensure comprehensive coverage of practice frameworks and technical specifications, we supplemented database searches with standards and grey literature (e.g., ISO/IEC 25012, ISO 8000) from relevant standards bodies and agencies.

2.3. Search Strategy

Search strings combined controlled vocabulary (where available) and free-text terms for the core concepts: data quality/information quality, dimensions, governance/ethics, and FAIR/standards.

PubMed (Title/Abstract + MeSH):

“data quality” [MeSH] OR “data quality” [tiab] OR “information quality” [tiab] AND (“Artificial Intelligence” [MeSH] OR “Artificial Intelligence” [tiab] OR AI [tiab]) AND (governance [tiab] OR ethics [tiab] OR FAIR [tiab])

Filters: 1996–2025; no language restriction. To align with the inclusion criteria (i) and focus the narrative synthesis, the search results were pre-filtered using the following “Article Types”: “Review”, “Systematic Review”, and “Meta-Analysis”.

Scopus (TITLE-ABS-KEY):

TITLE-ABS-KEY (“data quality” OR “information quality”) AND (“Artificial Intelligence” OR “AI”) AND (governance OR ethics OR FAIR) AND (PUBYEAR > 1995 AND PUBYEAR < 2026) Results were pre-filtered by “Document Type” to include “Review”.

To supplement the search and align with the inclusion criteria for conference papers and grey literature, the Scopus search was expanded. Using the same search string, an additional search was conducted and bulk-exported using the “Secondary documents” filter, which Scopus uses to categorize publications such as conference proceedings and book chapters. The “Preprint” (Beta) tab was also queried; however, as this section lacked a bulk export function, its results (N = 67) were manually screened by title for relevance. Potentially relevant preprints were then imported individually into EndNote for inclusion in the formal screening process.

Web of Science (TS Topic):

TS = (“data quality” OR “information quality”) AND TS = (“Artificial Intelligence” OR AI) AND TS = (governance OR ethics OR FAIR) Timespan: 1996–2025; Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI. To ensure methodological consistency with the

PubMed search, these results were pre-filtered using the “Document Type” filter for “Review Article”.

Google Scholar (broad discovery; first ~200 results screened):

(“data quality” OR “information quality”) AND (“Artificial Intelligence” OR “AI”) AND (governance OR ethics OR FAIR) Backward/forward citation chasing of sentinel papers and targeted searches on standards/agency sites supplemented all queries.

Search String Development and Validation

Search strings were developed iteratively through pilot searches and refined based on retrieved results and expert consultation among the author team. The search strategy combined three core concept blocks to ensure comprehensive coverage: data quality terminology, application domains and frameworks, and emerging themes such as artificial intelligence, governance, and FAIR principles. This multi-faceted approach allowed us to capture both foundational literature and contemporary developments.

For PubMed, we employed both Medical Subject Headings (MeSH) controlled vocabulary and free-text keywords to maximize sensitivity while maintaining specificity. The search combined terms related to data quality (“data quality” [MeSH] OR “data quality” [tiab] OR “information quality” [tiab]) with methodological and conceptual terms (“dimensions” [tiab] OR “framework” [tiab] OR “assessment” [tiab] OR “management” [tiab]), and domain-specific or thematic keywords (“healthcare” [tiab] OR “CRM” [tiab] OR “business” [tiab] OR “governance” [tiab] OR “FAIR” [tiab] OR “artificial intelligence” [tiab] OR “AI” [tiab]). Publication date filters were set to 1996–2025, and article type filters restricted results to Review, Systematic Review, and Meta-Analysis to align with our inclusion criteria and focus on synthesis papers rather than primary empirical studies.

For Scopus and Web of Science, we adapted the search strategy using platform-specific field tags. In Scopus, we used TITLE-ABS-KEY to search titles, abstracts, and keywords, whereas in Web of Science, we used the TS (Topic) field. To maintain consistency across databases, we applied document type filters, selecting “Review” in Scopus and “Review Article” in Web of Science. These filters ensured that our results were comparable and focused on synthesizing literature across all platforms.

Google Scholar searches employed simplified query strings due to platform limitations and the absence of advanced filtering options. We screened the first 200 results, ranked by relevance, a pragmatic approach commonly adopted in reviews that incorporate grey literature. Citation chasing—both backward (reference lists of key articles) and forward (articles citing seminal works)—supplemented database searches to identify foundational papers and recent applications that might not surface through keyword searches alone.

No language restrictions were applied to maximize coverage, though non-English articles were primarily identified through their English abstracts. Date filters (1996–2025) were consistently applied across all platforms.

2.4. Eligibility Criteria

Inclusion:

- (i) peer-reviewed articles, reviews, or conference papers that explicitly define, model, assess, or manage data quality;
- (ii) standards or technical reports describing data-quality frameworks or dimensions;
- (iii) sectoral applications (healthcare, business/CRM, public administration) illustrating implications for decision-making, governance, ethics, or FAIR/reproducibility.

Exclusion:

- (i) opinion/editorial pieces without methodological detail;

- (ii) works focused solely on cybersecurity or privacy without a substantive data-quality component;
- (iii) non-scholarly web content lacking sources;
- (iv) duplicates and superseded versions;
- (v) articles in which “data quality” was addressed only as routine data cleaning within the methodology of a specific empirical study (e.g., dietary intake and cancer association studies) without contributing to broader conceptual, methodological, or framework-oriented discussions of data quality.

2.5. Screening and Selection

All records were exported and deduplicated using EndNote version 20 and then exported to the app new.rayyan.ai. Screening proceeded in two stages: (1) title/abstract screening for relevance, followed by (2) full-text assessment to confirm eligibility. Screening was performed in duplicate (MGA, LGA); disagreements were resolved by discussion, with senior author adjudication (FGG or IAO) as needed.

2.6. Data Extraction

A piloted form captured bibliographic data, sector, and study type; definitions of data quality; dimensions and measurement approaches; governance/ethical components; alignment with FAIR/standards (ISO/IEC 25012, ISO 8000); tools/metrics; and reported outcomes/use-cases. Extraction was conducted independently by two reviewers with consensus adjudication.

To ensure the systematic and consistent extraction of information from included sources, we developed a structured data extraction form, which was piloted on 10 randomly selected articles and iteratively refined based on feedback from the extraction team. The form captured multiple categories of information: bibliographic details (authors, year, journal, DOI), study characteristics (domain, design, sample size or scope), definitions and conceptualizations of data quality, specific dimensions addressed (accuracy, completeness, timeliness, etc.), frameworks or standards referenced (such as ISO 8000, ISO/IEC 25012, or FAIR principles), methods used for data quality assessment or assurance, governance or ethical considerations discussed, and key findings and stated limitations.

Data extraction was performed in Microsoft Excel for Microsoft 365 Apps for enterprise, Version 2509 (Build 19231.20216), Current Channel (Microsoft Corporation, Redmond, WA, USA). One reviewer initially extracted each article, and a second reviewer independently verified it. Discrepancies or ambiguities were flagged for discussion and resolved through consensus or, when necessary, consultation with a senior author. This dual-extraction and verification process enhanced the reliability and completeness of extracted data.

For grey literature sources, including international standards documents (e.g., ISO 8000 series, ISO/IEC 25012) and technical reports from standards bodies or governmental agencies, we extracted normative definitions, prescribed dimensions and metrics, implementation guidance, and any case examples or sector-specific applications. We carefully recorded version numbers, publication years, and issuing organizations to ensure accurate citation and to distinguish between different editions or updates of the same standard.

2.7. Quality Appraisal and Source Credibility

Given the conceptual/standards-oriented scope, formal risk-of-bias tools for interventional/observational studies were not applicable. Instead, we appraised source credibility (peer review, issuing body for standards), recency, scope/transferability, and internal coherence of frameworks. For empirical applications, we documented design features germane to interpreting data-quality claims.

2.8. Synthesis Methods

Due to heterogeneity, we used thematic synthesis. Two reviewers independently coded texts, refined a shared codebook, and grouped findings into: (1) definitions/dimensions; (2) standards/formal frameworks; (3) governance and ethics; (4) FAIR and reproducibility; (5) sectoral applications; (6) assurance methods/tools. Themes were mapped to a conceptual schema that links data quality, governance/ethics, and FAIR principles to reproducibility and decision-making.

Software and Tools

Reference management and deduplication were performed using EndNote (Clarivate Analytics, Philadelphia, PA, USA) for initial import and automated duplicate detection, followed by manual verification in new.rayyan.ai. This process facilitated collaborative screening through shared libraries and manual tagging, and reasons for exclusion were documented for transparency and traceability.

Data extraction and synthesis were carried out using Microsoft Excel 365. Qualitative coding followed an iterative codebook approach: themes were identified from an initial subset of articles, refined through team discussions, and then systematically applied to the entire corpus. Codes were managed in parallel Excel workbooks by two independent reviewers (MGA, LGA), which were periodically merged and reconciled to resolve discrepancies and ensure uniform application of the coding framework.

We did not employ specialized systematic review software, such as Covidence or DistillerSR, as our narrative design did not require features specific to systematic reviews, including automated risk-of-bias assessment tools, meta-analytic capabilities, or standardized reporting templates for intervention studies. The selected tools provided sufficient functionality for the transparent and reproducible management of a narrative synthesis, while maintaining accessibility and cost-effectiveness for our research team.

2.9. Grey Literature and Standards Handling

Standards and technical reports (e.g., ISO 8000; ISO/IEC 25012) were treated as authoritative grey literature and included if they defined data-quality constructs or prescribed assessment/assurance requirements. Version and year, as well as scope, were recorded and cross-walked against academic frameworks.

2.10. Data Management and Reproducibility

The findings were synthesized thematically, highlighting dimensions, methodologies, and applications of data quality.

3. Definitions and Frameworks of Data Quality

3.1. Standards and ISO 8000

ISO 8000, the international data quality standard, originated in the NATO Codification System (NCS), a framework for cataloging material items with uniform descriptions and codes that improves inventory management and supply chain efficiency [16]. In 2004, NATO Allied Committee 135 and the Electronic Commerce Code Management Association (ECCMA) formalized its promotion as the basis for an international standard. This collaboration led to ISO 8000, which built on the NCS's emphasis on precise, standardized data descriptions and management practices, refined over decades [7,14,16].

ISO 8000 defines quality data as “portable data that meets stated requirements” [7]. Data must be separable from a software application (portable) and comply with explicit, unambiguous requirements set by the user or consumer [7]. The standard provides comprehensive frameworks for both conceptualizing and measuring data quality [8], thereby

aligning with broader initiatives in data portability and preservation, and enabling organizations to define, request, and verify data quality using standard conventions [8].

Complementing ISO 8000, ISO 25012 provides a foundational model with 15 dimensions of data quality, including accuracy, completeness, consistency, and traceability, initially designed for software but now widely applied across domains [9].

3.2. Frameworks

Wang and Strong developed a widely used framework with four dimensions: [15] (Table 1)

Table 1. Comparison of Major Data Quality Definitions and Frameworks (1996–2024). The table contrasts consumer-oriented conceptual models with formal international standards, highlighting key dimensions and practical applications across different organizational contexts.

Source/Standard	Definition of Data Quality	Key Dimensions/Emphasis	Notes
Wang & Strong [15]	Data quality is data that is fit for use by data consumers	Intrinsic (accuracy, objectivity) Contextual (relevance, timeliness, completeness) Representational (interpretability, consistency) Accessibility (access, security).	Highly influential conceptual framework; consumer-oriented.
Strong, Lee & Wang [14]	Emphasizes data quality as “fitness for use” in operations, decision-making, and planning.	Same four categories as above.	Extends the earlier framework with an organizational perspective.
Pipino, Lee & Wang [17]	Defines data quality through measurable attributes that reflect accuracy, completeness, consistency, and timeliness.	Quantitative measures for core dimensions.	Introduces practical tools for data quality assessment.
Ehrlinger & Wöß [18]	Data quality as a multidimensional construct is influenced by context and use.	Highlights timeliness, completeness, plausibility, integrity, and multifacetedness.	Extends beyond classical dimensions and focuses on big data.
Haug, Zachariassen [19]	Suggests that “perfect” data quality is neither achievable nor optimal; instead, the right level balances costs of maintenance vs. costs of poor data.	Trade-off between quality maintenance effort and business impact.	Cost-oriented perspective.

- Intrinsic—data should be accurate, reliable, and error-free;
- Contextual—relevance depends on the task or use case;
- Representational—clarity and proper formatting enable correct interpretation;
- Accessibility—data must be available to users when and where needed.

Data quality should be systematically assessed using both subjective perceptions and objective measures [17]. In healthcare, for instance, reliable data underpins patient records, diagnostics, research, and public health monitoring.

Information quality has been integrated into business process modeling to address potential deficiencies at the design stage, rather than relying solely on post-hoc cleansing [20]. In Customer Relationship Management (CRM), high-quality data enables segmentation, forecasting, and targeted marketing, directly affecting satisfaction and business performance, while poor-quality data undermines insights and decision-making [10,21]. Practical management requires preprocessing, profiling, and cleansing to ensure accurate analytics [1,11,17,18,22–24].

The dimensions based on ISO 25012 were later extended by aggregating 262 terms, a process that incorporated governance, usefulness, quantity, and semantics [25]. This

addresses terminology fragmentation and promotes cross-sector collaboration. A complementary approach grounded in the FAIR principles proposes quality ascertained from comprehensive metadata, decoupling intrinsic properties from subjective needs and facilitating reuse across contexts [25]. These principles have been extended to research software (FAIR4RS) [26].

Recent methods embed information quality requirements directly into business process models. One such technique involves “information quality fragments”: modular components designed to integrate checks for accessibility, completeness, accuracy, and consistency, which improves efficiency and reduces modeling time [27].

Beyond business contexts, frameworks for public administration emphasize both organizational and technological aspects. Studies of Ukrainian e-government reveal that staff skills, process design, and management practices are crucial for ensuring high-quality information in government [28].

A notable theoretical contribution is the “Objective Information Theory,” which defines information through a sextuple structure and nine metrics—such as extensity, richness, validity, and distribution—offering a systematic basis for quantitative analysis [29]. Building on this, research in the judicial domain has applied six-tuple information theory and rough set knowledge to develop measurable standards for document quality, showing that even text-driven contexts like court documents can benefit from formalized models that quantify accuracy, completeness, and semantic fidelity [30].

The FAIR principles have also been extended beyond data to research software, emphasizing that reproducibility depends not only on datasets but also on the tools and code that process them [31]. Together, these approaches underscore the diverse contributions of models of information quality to decision-making and broader outcomes across various sectors.

3.3. Scope of Data Quality in Data Analytics

In data analytics, data quality is paramount: analytical models and algorithms depend on accurate and complete inputs to generate meaningful insights and support sound decision-making. Models trained on flawed data can yield unreliable predictions, leading to misguided strategies. Conversely, high-quality data enhances the accuracy, fairness, robustness, and scalability of machine learning models, underscoring the importance of specialized tools in the era of data-centric AI [32,33].

Role in Data Analytics

Within Customer Relationship Management (CRM), data quality is especially critical. Reliable and consistent datasets enable accurate predictive modeling, such as customer lifetime value (CLV) estimations and RFM (Recency, Frequency, and Monetary) analysis, which supports effective segmentation, investment optimization, and customer equity assessment systems [13,32,34]. Well-maintained datasets enhance the effectiveness of predictive modeling. Poor data quality, by contrast, compromises analytics, leading to misinformed decisions, missed opportunities, and weakened strategies [32,35].

The importance of data quality tools is reflected in a 2022 systematic review, which identified 667 solutions for profiling, measurement, and automated monitoring [18]. Nevertheless, despite this breadth, many organizations still rely on rudimentary methods: a German survey found that 66% of companies validate data quality with Excel or Access, and 63% do so manually without long-term planning [36]. Within CRM, this gap between available technologies and organizational practices highlights the persistent challenges of implementing comprehensive, scalable data quality management.

4. Dimensions of Data Quality

4.1. Accuracy

Accuracy refers to the extent to which data represent the real-world construct they describe [35]. It can be quantified through error rates, field- and record-level measures [18,32], and newer metrics that distinguish between random and systematic errors [37]. Accuracy is especially critical for ensuring reliable insights in analytics and marketing, where decisions must reflect actual market conditions [12,21]. Accuracy is particularly challenging in CRM contexts, where customer attributes change frequently, requiring ongoing validation procedures to maintain reliable profiles [32].

4.2. Completeness

Completeness refers to the presence of all required data for a given purpose [18]. In CRM systems, completeness enables the creation of reliable customer profiles and supports accurate analytical outcomes, thereby reducing the risk of biased segmentation or decision-making [21].

4.3. Consistency

Consistency refers to the uniform representation of data across sources and systems, ensuring that values are not in conflict [34].

4.4. Timeliness

Timeliness concerns whether data are sufficiently up to date to support effective decision-making. Real-time data is vital in domains such as stock trading, where rapid changes demand instant responses. At the same time, daily or weekly updates may be sufficient for inventory management, helping to prevent overproduction or stockouts [38,39]. Timeliness depends on collection methods, processing speed, and update frequency, requiring organizations to minimize latency and ensure data is current [39–51].

A related challenge is data decay—the gradual obsolescence of information—seen in marketing, where up to 40% of email addresses change every two years [35]. Mitigating decay and maintaining timeliness are crucial for reliable analytics, resource efficiency, and improved decision-making [52].

4.5. Relevance

Relevance assesses how well data supports specific business needs and decision-making [15]. Data must serve its intended purpose, aligning with organizational objectives and challenges [53]. For example, accurate purchase histories may not fully explain customer churn if they lack information on satisfaction or service interactions; incorporating sentiment data provides more relevant insights.

Relevance is highly context-dependent, varying across different organizational areas [54]. Marketing departments prioritize demographics, purchasing behavior, and market trends to craft effective strategies. At the same time, financial teams focus on cash flow, revenue, and expenses for accurate planning. Thus, relevance is not static but shifts with users' goals and requirements.

Measuring relevance ensures that collected information contributes to informed decisions. Common approaches include tracking dataset usage, assessing “time to analysis” (shorter times reflect higher relevance), and gathering user feedback to identify gaps [55]. By aligning data with context-specific needs, organizations ensure that analytics are not only accurate but also meaningful for strategy and performance.

4.6. Validity

Validity ensures that data accurately represent real-world constructs and comply with predefined standards and business rules [56]. It assesses conformity to formats, ranges, and logical relationships—for example, a date of birth within a realistic range or an address following national conventions [54]. Validity encompasses several forms: content validity, which confirms that a dataset includes all relevant aspects [57]; construct validity, which ensures alignment with theoretical concepts [58,59]; and criterion-related validity, which evaluates the predictive power [60].

Invalid data can cause flawed analyses, poor decisions, and operational disruptions. Examples include failed deliveries to incorrect addresses [61], misleading financial data that affects investors, or inaccurate market research leading to wasted resources. Validity also supports integration and interoperability, as adherence to shared rules and formats facilitates combining data from multiple sources [57].

In the era of big data and AI, validity is increasingly critical: high-quality inputs underpin reliable models and algorithms, while advanced analytics can also help validate data [59,60]. Thus, validity is both a safeguard and an enabler of trustworthy data-driven processes.

4.7. Emerging Dimensions (Plausibility, Multifacetedness, Integrity)

Beyond classical attributes, new dimensions, such as governance, usefulness, quality, and semantics, have been proposed [62]. Governance refers to the accountability and authority structures that guide data activities. Usefulness emphasizes adaptability and reusability. Quantity addresses sufficiency and scalability. Semantics ensure that data meaning is preserved across systems. These additions reflect evolving understandings of data quality in contexts such as IoT, big data, and knowledge graphs.

Table 2 summarizes the main dimensions of data quality, including their definitions, practical examples, and measurement approaches.

Table 2. Core dimensions of data quality: operational definitions, practical examples, and measurement approaches. The report provides actionable criteria for assessing data quality across seven fundamental dimensions commonly applied in healthcare, business, and research contexts.

Dimension	Definition	Practical Example	Measurement Approach
Accuracy	The degree to which data correctly describes the real-world object or event.	Patient's recorded blood pressure matches the actual measurement.	Comparison against an authoritative source or ground truth.
Completeness	The extent to which all required data is present.	The customer database contains contact details for all clients.	Ratio of available values to required values; percentage of missing fields.
Consistency	Absence of contradictions within and across datasets.	A patient's birthdate is consistent across both electronic health records and insurance records.	Cross-field and cross-database validation checks.
Timeliness	The degree to which data is up to date and available when needed.	Stock market prices updated in real time.	Lag time between data generation and availability for use.
Validity	Degree to which data conforms to defined formats, rules, or ranges.	Postal codes follow the official national standard.	Validation rules, format checks, and range constraints.
Relevance	Appropriateness of data for the intended use.	Including clinical trial data when evaluating a new treatment.	Expert judgment; alignment with analytical or decision-making needs.
Uniqueness	The degree to which data is free of duplicate records.	Each patient has a single unique medical record number.	Duplicate detection and record linkage algorithms.

5. Impact and Evidence: Consequences and Case Studies of Data Quality Management

5.1. Multidimensional Consequences of Poor Data Quality

Organizations that neglect data quality face serious and multifaceted consequences. Research demonstrates that poor data quality generates quantifiable harm across multiple dimensions: financial losses, compromised decision-making, operational inefficiency, reputational damage, and missed strategic opportunities.

5.1.1. Financial Impact

The financial consequences of poor data quality are substantial and well-documented. Gartner estimates annual losses of USD 12.9 million per organization, while IBM reports that U.S. businesses collectively lose over USD 3.1 trillion annually due to data quality issues [61,63]. Industry-specific studies indicate that organizations can lose up to 6% of their sales due to poor customer data management]. In CRM systems, inaccurate or inconsistent records distort insights, inflate costs, and reduce efficiency. Such issues often require costly cleansing and manual intervention—costs that exponentially exceed those of prevention [64–68].

5.1.2. Operational and Strategic Consequences

Low-quality data leads directly to flawed segmentation, poor targeting, and reduced strategic effectiveness. In CRM systems, outdated contact details or incorrect names undermine marketing efforts and customer relationship strategies. Time, money, and credibility are wasted addressing the consequences of these errors [69–75]. Industry experts estimate that data quality challenges cost companies 10–30% of their annual revenue [55].

5.1.3. Resource Waste and Operational Inefficiency

Cleaning inaccurate or incomplete data is a significant and resource-intensive task [76]. Manual and automated correction efforts are expensive and often consume disproportionate organizational resources—tasks performed with incorrect data cost 100 times more than those performed with accurate data [61,77]. In CRM systems, poor validation can erode user trust and compromise adoption, particularly in small and medium-sized enterprises (SMEs), where frequent inconsistencies are typical [78–81]. Robust cleaning practices—such as removing duplicates, filling in missing values, and discarding outdated entries—reduce remediation costs and risks; yet, many organizations continue to rely on rudimentary methods [18]. Similar findings have been reported in the CRM literature, which highlights the persistence of low-maturity data quality practices despite the recognized strategic importance of clean data [10].

5.1.4. Reputational Damage and Stakeholder Trust

Incorrect interactions or failed campaigns erode customer trust and damage brand image [82]. When organizations deliver poor customer experiences due to inaccurate data—such as duplicate mailings, incorrect product recommendations, or failed service interactions—the reputational consequences extend beyond immediate financial losses [83–85]. Trust, once damaged, is difficult to restore and has a direct impact on customer loyalty and organizational competitiveness [85–89].

5.1.5. Missed Opportunities and Competitive Disadvantage

Poor-quality data contributes to wasted potential and missed strategic opportunities. Up to 45% of generated leads are deemed unusable due to duplication, invalid email addresses, or incomplete records [90]. In healthcare, incomplete patient records can obscure

critical patterns, delaying medical discoveries and compromising research progress [91–93]. In business intelligence, poor-quality data hinders accurate market segmentation and opportunity identification, preventing the identification of opportunities and hindering innovation and competitive positioning [94]. Such missed opportunities amplify long-term competitive disadvantages, as organizations lose market share to competitors with superior data practices [95]. Table 3 summarizes these multidimensional consequences across sectors, demonstrating the financial, operational, reputational, and strategic impacts of inadequate data quality management.

Table 3. Multidimensional consequences of poor data quality across sectors. Seven major consequence categories are illustrated with real-world examples from healthcare, finance, logistics, and public administration, demonstrating the financial, operational, reputational, and regulatory impacts of inadequate data quality management.

Consequence	Definition	Practical Example	Impact/Cost
Faulty decision-making	Wrong or suboptimal choices based on inaccurate data.	A hospital prescribes inappropriate treatment due to errors in lab data.	Patient harm, liability risks, loss of trust.
Financial losses	Direct or indirect costs from incorrect, incomplete, or duplicated data.	A bank suffers multimillion-dollar losses due to flawed credit risk models.	Wasted resources, loss of revenue.
Operational inefficiencies	Processes slowed or disrupted due to unreliable information.	Logistics companies misroute deliveries due to inaccurate addresses.	Increased workload, delays, and higher costs.
Reputational damage	Erosion of trust from stakeholders, customers, or the public.	Data breaches and reporting errors damage a company’s brand.	Customer attrition, lower market share.
Regulatory and legal risks	Non-compliance with laws and standards due to poor data.	A pharmaceutical firm fails an audit due to inconsistent records.	Fines, sanctions, reputational harm.
Missed opportunities	Failure to identify insights or innovations.	Retailer loses potential sales due to incomplete CRM data.	Reduced competitiveness, slower growth.
Misleading analytics	Models or reports based on flawed inputs lead to invalid results.	Overestimation of flu outbreaks by Google Flu Trends.	Misallocation of resources leads to a loss of credibility.

Table 3 summarizes key consequences across domains.

5.2. High-Profile Failures: Learning from Data Quality Disasters

Real-world cases vividly illustrate the stakes of data quality and the serious consequences of inadequate governance and technical oversight. Across various sectors, including aerospace, finance, healthcare, technology, and public administration, data quality failures have led to catastrophic losses, operational disruptions, and even loss of life.

5.2.1. Aerospace: NASA’s Mars Climate Orbiter

One of the most striking examples of the consequences of poor data quality occurred in space exploration. NASA’s Mars Climate Orbiter was destroyed by a mismatch between imperial and metric units—a \$125 million failure of consistency [96–100]. This case exemplifies how inadequate data validation processes and insufficient governance can lead to catastrophic losses. The failure underscores that data quality is not merely a techni-

cal concern; it is a mission-critical issue that requires robust verification across multiple organizational levels [101].

Lesson: Standardization, validation, and clear governance protocols are non-negotiable in high-stakes domains.

5.2.2. Technology: Unity Technologies' Audience Pinpoint Error

A striking corporate example is Unity Technologies' Audience Pinpoint algorithmic failure, in which poor input data fed into machine learning models produced faulty algorithms that ultimately resulted in USD 110 million in losses and a 37% decline in stock value [102]. The error illustrates how inadequate data quality upstream—at the data sourcing and preparation stage—can propagate through AI systems, leading to significant business failures [103].

Lesson: AI systems are only as reliable as their training data; governance must begin at the data ingestion stage.

5.2.3. Public Administration: Data Inaccuracy in Criminal Justice

Research into computerized criminal records in the U.S. reveals that these records are often inaccurate or incomplete, thereby undermining the reliability of the justice system [104–106]. Poor information quality introduces inefficiencies and errors with serious social consequences highlighting the need to align process design with information system requirements from the outset [107–113]. This case demonstrates that even critical public-sector systems remain vulnerable to data quality failures [112].

Lesson: Organizational accountability and robust process design are essential, even in mission-critical public functions.

5.2.4. Financial Sector: Data Silo Architecture and Systemic Risk

In financial institutions, legacy “data silos” can perpetuate duplication, inconsistency, and limited access, hampering analytics and decision-making [111,113–116]. The decentralized storage and inconsistent data entry practices created systemic risk throughout the organization [117,118]. Transitioning toward modern paradigms, such as the Data Mesh architecture, has improved availability and quality across the enterprise, demonstrating that organizational architecture directly impacts data quality outcomes [119–122].

Lesson: Data governance is as much an architectural issue as a technical one; silos are a structural vulnerability [119,123–125].

5.3. Success Stories: Data Quality as a Strategic Asset

In contrast to failures, organizations and sectors that invest in robust data quality frameworks achieve measurable competitive advantages, improved decision-making, and strategic success.

5.3.1. E-Commerce: Netflix's Data-Driven Success

Netflix's personalized recommendation system illustrates how high-quality data can become a strategic asset. By leveraging user behavior data with rigorous quality standards, Netflix drives over 80% of platform consumption through recommendations. The system's reliability depends on continuous data validation, user feedback integration, and quality monitoring—not just advanced algorithms [126–130].

Lesson: High-quality data, coupled with effective governance, enables competitive differentiation and fosters customer loyalty.

5.3.2. Healthcare: Multicenter Clinical Database Excellence

A multicenter clinical database study has demonstrated the effectiveness of semi-automated data quality assessments in detecting widespread issues across multiple healthcare facilities [131–133]. A balanced approach that integrates automation with targeted manual review optimizes both efficiency and accuracy [134,135]. Organizations implementing comprehensive quality frameworks report greater data reliability, fewer clinical errors, and improved patient outcomes [136–140].

Lesson: In healthcare, robust data governance is essential for protecting patient safety and maintaining research integrity; investment is non-negotiable.

5.3.3. Business Operations: CRM Implementation with Strong Governance

Organizations that implement comprehensive CRM governance frameworks report measurably improved customer segmentation, more effective targeted marketing, stronger customer relationships, and higher operational efficiency. These successes share common traits: clear ownership, standardized collection, automated validation, regular monitoring, and sustained leadership commitment to continuous improvement [141–144].

Lesson: Sustained success requires embedding quality mechanisms into workflows, not retrofitting them afterward.

5.4. Cross-Sector Patterns: Where Data Quality Determines Outcomes

Synthesizing failures and successes across sectors reveals consistent patterns about where data quality has the highest organizational impact and which governance approaches work most effectively.

5.4.1. Healthcare Sector

Healthcare demonstrates perhaps the highest stakes for data quality. Patient safety, clinical decision-making, research integrity, and public health monitoring all depend on reliable data [145,146]. Qualitative research in Ethiopian public health facilities reveals that inadequate training, weak governance, and insufficient infrastructure compromise the timeliness, completeness, and accuracy of care, directly affecting patient outcomes [147]. Conversely, organizations that implement the WHO's Data Quality Review toolkit and establish clear governance frameworks achieve measurably better outcomes [136].

5.4.2. Finance and Investment

Financial institutions face unique data quality challenges due to the volume, velocity, and regulatory requirements of economic data. Successful organizations transition to modern architectures that prioritize data accessibility, consistency, and real-time monitoring, reducing decision latency and regulatory risk [145–147].

5.4.3. E-Commerce and CRM

The customer relationship sector demonstrates that data quality directly correlates with customer satisfaction, retention, and revenue [10,148–151]. Poor CRM data quality leads to failed campaigns, customer frustration, and a loss of market share [152–154]. Successful implementations combine standardized data entry, automated validation, continuous monitoring, and user feedback integration, creating virtuous cycles of improvement [7,17,23,155–157].

5.4.4. Public Administration

Government sector research shows that information technology alone does not ensure data quality [158,159]. Persistent shortcomings in public administration, including insufficient staff training, weak governance structures, and organizational silos, undermine

data reliability [160,161]. Successful e-government initiatives combine clear governance, adequate training, process redesign, and technological solutions [159,162].

5.4.5. Small and Medium Enterprises (SMEs)

SMEs remain particularly vulnerable to data quality issues [23]. A study of 85 UK SMEs found that deficits in technical skills and resources hinder the adoption of robust practices [163]. However, SMEs that prioritize data quality frameworks and invest in staff training achieve a competitive advantage in their markets [23,164].

Table 4 presents a comparative synthesis of failures and successes across these sectors, illustrating both the variability of consequences and the universality of core governance principles.

Table 4. Comparative case studies of data quality failures and successes. High-profile failures and successful implementations are contrasted across multiple sectors, illustrating how data quality has a direct impact on financial performance, operational efficiency, public trust, and organizational competitiveness.

Case/Organization	Domain	Data Quality Issue	Consequence
Failures			
Equifax [165]	Finance/Credit reporting	Inaccurate and poorly managed consumer credit data [166]	Erosion of public trust; legal and financial consequences [167]
NASA Mars Climate Orbiter [101]	Aerospace/Engineering	Unit mismatch (imperial vs. metric) not reconciled in data systems [168]	Spacecraft loss (~\$125 million)
Mid-sized enterprise (CRM migration) [35]	Business/CRM	Data quality challenges during migration from legacy systems [23,64,169,170]	Errors, inconsistent formats, and disruption in customer management
Large home appliance business [21]	Retail/CRM	Low completeness, timeliness, and accuracy of customer data [21,171]	Ineffective campaigns, reduced loyalty, and weak predictive performance
University fundraising CRM [32].	Education/Fundraising	Outdated, incomplete, and inaccurate alumni data [172,173]	Reduced donor identification, inefficient fundraising, wasted resources
Target [126].	Retail/CRM	Predictive analytics revealed sensitive customer information [174]	Public backlash over privacy intrusion
Google Flu Trends (2008–2013) [175–177].	Public health analytics	Overfitting and reliance on biased signals	Overestimation of flu cases; credibility loss [178]
Amsterdam Tax Office [35].	Public sector	Duplicate and inconsistent taxpayer records [179–181]	Inefficient operations; reduced compliance [182]
Healthcare organizations [34,183,184]	Healthcare/CRM	Incomplete or inconsistent patient data in electronic health records [131,185]	Medical errors, patient safety risks
Successes of Data Quality			
Netflix Recommendation System [126–130].	Entertainment/Business	Leveraging high-quality behavioral data for personalization [127]	Recommendations drive 80% of content consumption; increased engagement and revenue
Freight forwarding industry [20]	Logistics/Freight forwarding	Workflow-embedded quality checks across logistics processes [186]	Improved coordination, fewer customs delays, and reduced correction costs

5.5. Synthesis: Why Data Quality Governance Matters

These cases and patterns collectively demonstrate that data quality failures are not isolated technical incidents but symptoms of systemic governance gaps. Poor data quality consistently arises from:

- Inadequate process design at the point of data collection
- Insufficient validation and verification mechanisms
- Organizational silos that prevent information sharing and consistency
- Weak leadership commitment to quality as a strategic priority
- Insufficient training and unclear accountability structures
- Delayed or absent monitoring and maintenance processes

Conversely, successful organizations share these characteristics:

- Clear governance structures with defined roles and accountability
- Quality mechanisms embedded throughout the data lifecycle (not retrofitted)
- Continuous monitoring and feedback loops
- Leadership commitment and resource allocation
- Regular training and stakeholder engagement
- Alignment of technology with organizational processes and objectives

These patterns underscore a critical insight: data quality challenges are as much organizational and social as they are technical [187–189]. Technology alone—even sophisticated AI-driven validation systems—cannot resolve governance and cultural barriers [190,191]. Sustainable data quality requires the integration of technical solutions, organizational design, clear governance, and committed leadership [192].

The financial, operational, and strategic costs of poor data quality far exceed the investments required to prevent and maintain continuous quality assurance [191,193]. For practitioners and organizations, the decision is clear: treating data as a strategic asset and investing in comprehensive quality governance is not optional but essential for organizational success, scientific reproducibility, and informed decision-making [187,189].

6. Ensuring Data Quality: Methods and Best Practices

Ensuring high data quality requires a combination of standardized methodologies and continuous validation. Effective data collection is fundamental: standardized procedures and tools minimize errors at the source, while validation checks confirm that data entries meet predefined criteria [194].

6.1. Data Collection Practices

In CRM adoption among SMEs, robust data collection processes are crucial for capturing accurate and relevant customer information. This involves integrating multiple internal and external sources and ensuring that data is timely and aligned with business objectives. Systematic frameworks and staff training help maintain consistency and reduce errors such as duplicate or incomplete records, which can undermine CRM performance [12].

Best practices include the use of automated data entry systems, regular cleaning processes to eliminate inaccuracies, and validation checks to ensure consistency and accuracy. Continuous monitoring and maintenance are crucial as data sources and systems evolve, ensuring that quality standards are consistently maintained [10].

6.2. Data Cleaning and Validation

Data cleaning and validation are crucial for maintaining data quality, as they involve detecting and correcting inaccuracies, inconsistencies, and missing values [12]. Automated tools can efficiently flag anomalies, but critical production environments often require a

balanced approach that integrates manual review to ensure accuracy and reliability. A multicenter clinical database study demonstrated the effectiveness of semi-automated assessments in detecting widespread issues, while minimal manual checks remained necessary to address errors missed by automation [195]. Regular monitoring and quality reviews help sustain reliable datasets [12,196]. Establishing metrics and periodic audits ensures data remains accurate, timely, and fit for analysis. In CRM systems, poor validation can erode user trust and compromise adoption [34,35], particularly in SMEs facing frequent inconsistencies [35]. Preventive controls and improved entry processes further minimize errors [13].

Advanced methods, including deep learning models such as the Denoising Self-Attention Network (DSAN), enhance the imputation of mixed-type tabular data by leveraging self-attention and multitask learning to predict substitutes for both numerical and categorical fields [197]. Data fusion also improves reliability by integrating heterogeneous sources, thereby reducing uncertainty and providing more comprehensive representations [198].

Ultimately, rigorous validation and governance policies [1] combined with automated tools and audits [12] ensure consistent data quality. This is crucial for CRM, where reliable data directly impacts campaign success, customer loyalty, and business performance [21].

6.3. Monitoring and Maintenance

Continuous monitoring and regular maintenance preserve data integrity by detecting anomalies, cleansing outdated records, and ensuring consistent definitions across systems [18,199–202]. Effective collaboration among IT teams, data stewards, and business users sustains reliability throughout the data lifecycle [203–205].

6.4. Governance/Ethics

Effective governance requires clear roles, responsibilities, and leadership commitment throughout the data lifecycle [206], including compliance with regulations such as GDPR.

An integrated data and AI governance model spans four phases: (1) data sourcing and preparation with ethical acquisition and bias detection, (2) model development with fairness-aware validation, (3) deployment and operations with drift monitoring, and (4) feedback and iteration to update policies dynamically [207]. Automated checkpoints across these phases link data integrity with model ethics.

Responsible governance must extend beyond compliance to embed fairness, accountability, and inclusivity [208]. Ethical lapses can erode trust, reinforce biases, and lead to harmful outcomes across various domains, such as healthcare and finance [209]. Embedding ethical principles within governance frameworks ensures data quality is evaluated not only for efficiency but also for societal impact.

These concerns are amplified in large language models, where massive training datasets raise challenges of provenance, bias, and misuse. Robust governance of LLMs is therefore essential to prevent social and economic harm [210].

7. Challenges and Proposed Solutions

Managing data quality in CRM systems presents challenges, including decentralized storage, inconsistent data entry, and limited integration of diverse data sources. Proposed solutions include standardizing collection procedures, using advanced integration tools, and implementing comprehensive management frameworks to strengthen accuracy and consistency [10].

7.1. Technical Challenges

Technical obstacles stem from data heterogeneity, scale, and limited automation. Integrating multiple sources and processing large volumes of inconsistent or outdated information remain critical difficulties [21]. These issues are exacerbated by web-based platforms and SaaS models, which aggregate data from various internal and external sources [24]. Many organizations continue to rely on outdated or manual quality checks, which are inadequate given the volume and velocity of contemporary data [18].

A further barrier is the lack of standardized terminology across sectors; Miller identified 262 distinct terms in the literature to describe aspects of data quality, which complicates interoperability and hinders the adoption of automated assurance tools [62]. Addressing these challenges requires not only technical solutions but also harmonized frameworks that enable consistency and support effective, data-driven decision-making.

7.2. Organizational Challenges

The significance of data quality extends beyond technical systems to organizational processes, as illustrated by studies showing that even computerized criminal records in the U.S. are often inaccurate or incomplete [211]. Poor information quality undermines workflows by introducing inefficiencies and errors, highlighting the need to align process design with information system requirements from the outset [20]. Common issues include inadequate data entry, poorly designed systems [13], weak leadership commitment, and insufficient resource allocation.

AI adoption introduces new challenges. Automation of complex tasks can deskill experts, reducing their role to passive monitoring and eroding the valuable expertise they possess. The autonomy of AI systems also creates ambiguity over accountability, raising legal and ethical questions when errors occur [208].

A systematic review of Data Science projects identified 27 barriers grouped into six clusters: people, data and technology, management, economic, project, and external factors. The most frequently cited included insufficient skills, lack of management support, poor data quality, and misalignment between project goals and company objectives [212]. Addressing these barriers requires comprehensive governance, continuous training, and systematic frameworks for data handling and interpretation [34].

Perfect data quality is neither achievable nor cost-effective. Organizations should instead pursue the “right level” of quality, balancing maintenance efforts against the costs of poor quality [213]. For large datasets, the marginal benefits of further improvements diminish over time [32]. Relevance is another concern: as business needs evolve, datasets may lose strategic value, and silos that isolate data within departments hinder efficient decision-making [214,215]. Integrating quality assurance into process design reduces errors and enhances reliability [20,27].

Public administration research confirms that technology alone does not ensure high-quality information. Shortcomings such as insufficient training, weak governance, and organizational inefficiencies limit outcomes, underscoring that data quality challenges in government are rooted as much in people and structures as in systems [28].

7.3. Solutions

Organizations can address data quality challenges by adopting advanced, automated tools, including AI and machine learning. These technologies enable continuous monitoring, anomaly detection, and large-scale validation, surpassing the capabilities of manual methods. Recent work demonstrates the potential of multi-agent AI systems, in which “planner” and “executor” agents automatically generate and execute PySpark validation code from natural-language prompts, adapting to evolving data sources [216]. However,

user acceptance remains a barrier: a survey of an AI-generated e-commerce platform (ChatGPT-4 and DALL·E 3) revealed concerns about security and trust despite positive evaluations of functionality and aesthetics [217].

Effective solutions also require developing business-relevant quality metrics to guide decision-making and improvement strategies. AI-based frameworks can support Business Intelligence (BI) and Decision Support Systems (DSS) through layered architectures: data ingestion, quality assessment (e.g., Isolation Forests, BERT), cleansing and correction (e.g., KNN, LSTM), and monitoring with feedback loops [218]. Predictive approaches, such as XGBoost, can correct anomalies across multiple dimensions (accuracy, completeness, conformity, uniqueness, consistency, and readability), improving composite Data Quality Index scores from 70% to over 90% in large-scale applications [219].

Beyond technology, sustainable quality depends on robust governance. Frameworks should define criteria for data relevance, ensure alignment with organizational objectives, and include regular audits to maintain validity [56]. Engaging stakeholders in data management further strengthens the practical utility of collected information.

8. Data Quality in the AI Era: Paradigm Shifts and Governance Implications

The emergence of artificial intelligence and machine learning as dominant paradigms for data-driven decision-making has fundamentally altered the landscape of data quality [94,220,221]. Traditional frameworks, developed in an era when data primarily served human analysts [222], must now accommodate systems in which algorithms themselves are the primary consumers of data [223–225]. This section examines how AI reshapes the conceptualization of data quality, challenges conventional governance structures, and necessitates the integration of ethical considerations as core quality dimensions.

8.1. How AI Reshapes the Definition of Data Quality

Traditional definitions of data quality center on the concept of “fitness for use,” whereby data are evaluated in relation to their capacity to support human decision-making processes [1,3,226]. However, the rise of AI introduces a fundamental shift: data must now be fit not only for human interpretation but also for algorithmic processing and training machine learning models. This dual requirement expands the scope of data quality beyond conventional dimensions.

Such failures demonstrate that poor input data can propagate through AI systems, resulting in significant financial and operational consequences [165].

AI systems impose specific requirements that extend beyond traditional quality dimensions.

AI-Readiness: Data must be structured, annotated, and formatted in a way that allows algorithms to process it efficiently. This includes requirements for standardized schemas, consistent labeling in supervised learning contexts, and appropriate representation of features for model input.

Algorithmic Fairness: Data quality must explicitly account for representation and balance across demographic groups, domains, or categories to ensure fairness. Biased training data perpetuate and amplify discrimination in algorithmic decision-making, as demonstrated in healthcare, criminal justice, and financial domains [200,206,208,227–236].

Model Explainability and Traceability: High-quality data must support interpretability requirements. Provenance, lineage, and metadata become critical not merely for reproducibility but for understanding how algorithmic decisions are derived—essential for regulatory compliance and ethical accountability.

These AI-specific requirements challenge the traditional separation between intrinsic data properties (accuracy, completeness, and consistency) and contextual dimensions (relevance and timeliness). In AI contexts, what constitutes “accurate” or “complete” data

depends fundamentally on the model architecture, training objectives, and deployment context. A dataset may be adequate for descriptive analytics yet wholly insufficient for predictive modeling if it lacks temporal coverage, feature diversity, or a sufficiently large sample size for robust generalization.

Furthermore, AI introduces dynamic quality requirements. Unlike static reporting or analysis, where data quality can be assessed once and maintained, machine learning models require continuous monitoring of data quality throughout their operational lifecycle. Model drift—the degradation of predictive performance over time as real-world conditions change—requires ongoing validation to ensure that training data remain representative of current conditions [18,63,69,181–183,205,207]. This temporal dimension transforms data quality from a one-time assessment into a continuous assurance process.

Recent adaptations of the FAIR framework emphasize AI readiness and the need for rigorous attention to provenance, bias detection, and ethical sourcing [25,26,208,236–240]. However, achieving AI readiness requires more than technical interoperability; it demands rigorous attention to data provenance, bias detection, and ethical sourcing—considerations that extend far beyond the original FAIR framework’s scope [236,241–245].

8.2. AI-Driven Challenges to Traditional Governance Frameworks

International standards such as ISO 8000 and ISO/IEC 25012 provide foundational frameworks for data quality management [6,7,246]. These standards codify dimensions (accuracy, completeness, consistency, timeliness, validity) and prescribe assessment methodologies that have proven effective in traditional business intelligence, reporting, and operational contexts. However, these frameworks were developed in an era preceding widespread AI adoption and therefore do not fully address the unique governance challenges introduced by algorithmic decision-making systems.

The regulatory landscape for AI governance is rapidly evolving to address these challenges through comprehensive frameworks that mandate systematic approaches to AI risk management and quality assurance [247]. Three pivotal normative instruments exemplify this transformation: ISO/IEC 42001:2023 [246] establishes the first international management system standard specifically for AI, providing organizations with a structured framework for the responsible deployment of AI that integrates quality management principles with AI-specific risk controls [248]. The European Union’s AI Act (Regulation 2024/1689) [249] introduces the world’s first comprehensive regulatory framework with legally binding requirements for high-risk AI systems, mandating rigorous data governance, transparency measures, and human oversight mechanisms that fundamentally reshape how organizations must approach data quality throughout the AI lifecycle [250]. Similarly, NIST’s AI Risk Management Framework (AI RMF 1.0) [251] offers a voluntary yet influential framework that emphasizes trustworthy AI characteristics—including validity, reliability, accuracy, and fairness—directly linking data quality dimensions to AI governance outcomes. Together, these frameworks signal a paradigm shift from voluntary best practices to mandatory compliance requirements, compelling organizations to implement formal AI governance structures that embed data quality assurance as a core operational imperative, rather than a technical consideration [252]. This regulatory convergence creates both challenges and opportunities: while organizations face increased compliance burdens and documentation requirements, these frameworks also provide standardized methodologies for addressing the complex interplay between data quality, algorithmic fairness, and operational excellence in AI systems [253]. The implications extend beyond compliance, as these normative instruments are reshaping market dynamics, with organizations that demonstrate robust AI governance and data quality practices gaining competitive advantages in terms of trust, market access, and regulatory alignment [254,255].

Scale and Velocity: AI systems often process data at volumes and speeds that render manual validation infeasible. Traditional quality assurance methods—such as manual review, rule-based validation, and periodic audits—cannot keep pace with the continuous data ingestion and real-time processing demands of production AI systems, and use approaches wholly inadequate for AI-driven environments [218,256].

Opacity and Interpretability: Machine learning models, particularly deep neural networks, operate as “black boxes” whose internal decision logic is opaque even to their developers. This opacity creates accountability gaps: when an AI system produces an erroneous or harmful output, determining whether the failure stems from poor data quality, algorithmic bias, model misconfiguration, or deployment context becomes exceedingly difficult [113,207,208,257]. Traditional governance frameworks assume that quality issues can be traced to identifiable data sources and corrected through targeted interventions—an assumption that breaks down in complex AI pipelines. Automation introduces accountability challenges, as decision-making shifts from humans to algorithms, requiring precise governance mechanisms to assign responsibility [208,209,237].

Bias Amplification: Traditional data quality frameworks emphasize representational accuracy—whether data correctly reflect real-world states. However, AI systems can perpetuate and amplify biases present in training data, even when those data are “accurate” in a technical sense. For instance, historical hiring data may accurately reflect past decisions but encode discriminatory patterns that AI recruitment tools then reproduce at scale [228,229]. This challenge requires governance frameworks to evaluate not only factual accuracy but also fairness, equity, and social impact—dimensions that are often absent from conventional standards.

Temporal Dynamics and Model Drift: Unlike static datasets used for one-time analysis, AI models operate in production environments where data distributions shift over time. Seasonal patterns, market changes, demographic shifts, or external shocks (e.g., pandemics, regulatory changes) can render training data unrepresentative, degrading model performance [258]. Governance frameworks must therefore incorporate continuous monitoring, drift detection, and retraining protocols—requirements not emphasized in traditional standards focused on periodic assessment.

Integrated governance frameworks link ethical acquisition, model validation, continuous monitoring, and adaptive feedback throughout the AI lifecycle [207].

However, implementing such integrated frameworks faces significant organizational barriers. A systematic review of data science projects identified 27 barriers grouped into six clusters: people, data and technology, management, economic, project, and external factors [212]. The most frequently cited included insufficient skills, lack of management support, poor data quality, and misalignment between project goals and organizational objectives. These findings show that governance failures are as much organizational and cultural as they are technical—a reality that technical standards alone cannot resolve.

8.3. AI Ethics as an Integral Component of Data Quality

The integration of ethical considerations into data quality frameworks represents one of the most significant paradigm shifts in the era of AI. Traditional data quality management focused primarily on technical accuracy and operational efficiency. In contrast, contemporary AI governance recognizes that data quality cannot be divorced from ethical principles such as fairness, accountability, transparency, and equity [227,233,240,259].

Bias and Fairness: Biased training data lead directly to biased algorithms, which in turn produce discriminatory outcomes across healthcare, criminal justice, lending, hiring, and other high-stakes domains [228,229,260]. The concept of “fairness” in AI contexts is multifaceted, encompassing statistical parity (equal outcomes across groups), equal

opportunity (equal true positive rates), and individual fairness (treating similar individuals similarly). Achieving any of these fairness criteria depends fundamentally on the quality, representativeness, and balance of training data. Consequently, assessing and mitigating bias is no longer an optional ethical enhancement but a core data quality requirement.

Accountability and Transparency: When AI systems make consequential decisions—approving loans, diagnosing diseases, recommending sentences—stakeholders have a right to understand how those decisions were reached. Data provenance, lineage, and quality documentation become essential for accountability [257,259]. If a model produces a harmful outcome, investigators must be able to trace the decision back through the algorithmic pipeline to the underlying data, assess whether quality issues contributed to the failure, and assign responsibility appropriately. Governance frameworks must therefore mandate comprehensive metadata, audit trails, and transparency mechanisms—requirements that extend well beyond traditional data quality assurance.

Privacy and Consent: Large-scale AI systems, particularly large language models (LLMs), are trained on massive datasets often scraped from public sources without explicit consent [227]. Data quality governance in this context must address not only technical accuracy but also ethical sourcing, privacy preservation, and respect for data subjects' rights. Technologies such as federated learning and homomorphic encryption enable training on distributed or encrypted data, embedding privacy-by-design into the technical architecture [261]. These approaches demonstrate how governance is being increasingly integrated directly into AI systems, rather than being treated as an external compliance layer.

Hallucinations and Misinformation: LLMs exemplify the risks of poor data quality in AI contexts. Hallucinations—confident but factually incorrect outputs—arise from inadequate governance during training and evaluation, often resulting from misuse, weak security, or poor-quality inputs [262–267]. Misinformation, once generated by an AI system, can propagate rapidly and at scale, causing social harm that far exceeds the original data quality failure [268,269]. Governance frameworks must therefore integrate ethical safeguards—such as adversarial testing, red-teaming, and continuous output monitoring—to sustain trust and prevent harm [251].

Emerging governance frameworks explicitly incorporate these ethical dimensions. For instance, fairness-aware machine learning techniques integrate bias detection and mitigation directly into model training pipelines, using techniques such as adversarial debiasing, reweighting, and fairness constraints [270]. Similarly, explainable AI (XAI) methods provide interpretability mechanisms that enable stakeholders to interrogate model decisions and identify data quality issues that contribute to erroneous outputs [271]. These technical innovations represent a convergence of data quality assurance and ethical AI governance, reflecting the recognition that technical and ethical dimensions are inseparable.

Figure 1 illustrates these interactions, showing governance as a mediator between technical tools, organizational practices, and ethical safeguards. Large-language-model behavior further highlights how data quality and governance jointly determine reliability and trust [23,210,272]. In this integrated view, data quality is not merely a technical prerequisite for AI but a sociotechnical construct that encompasses accuracy, fairness, transparency, accountability, and societal impact.

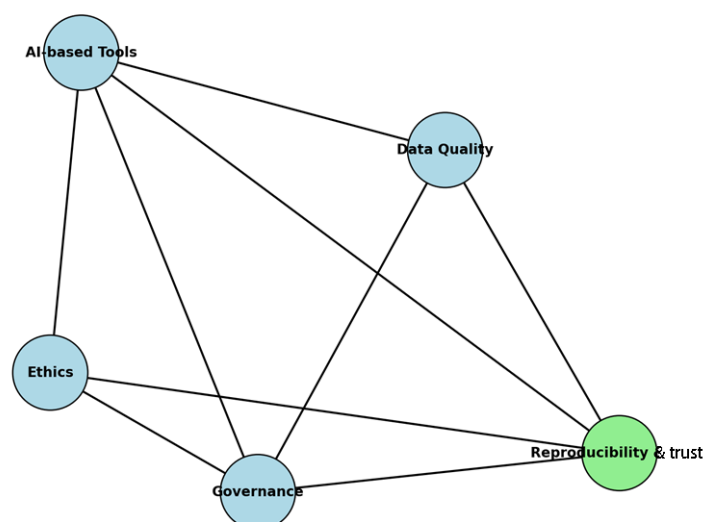


Figure 1. Conceptual framework integrating data quality, governance, ethics, and AI tools for reproducibility and trust. Light blue nodes represent core data quality management components; the green node highlights reproducibility and trust as the collective outcome. Bidirectional arrows indicate mutually reinforcing relationships, demonstrating that technical solutions alone are insufficient without robust governance and ethical safeguards.

8.4. From “AI for Data Quality” to “Data Quality for AI”: A Bidirectional Relationship

The relationship between AI and data quality is bidirectional: AI systems both depend on high-quality data and offer powerful tools for assuring data quality at scale. However, current practices and market offerings remain imbalanced, with greater emphasis on “data quality for AI” (ensuring inputs to models meet quality standards) than on “AI for data quality” (using AI to detect and correct quality issues) [22].

AI for Data Quality: Advanced automated tools, including machine learning algorithms, enable continuous monitoring, anomaly detection, and large-scale validation that surpasses manual methods [113,154,184,273]. Multi-agent AI systems demonstrate the potential for full automation, with “planner” and “executor” agents automatically generating and executing PySpark validation code from natural-language prompts, adapting dynamically to evolving data sources.

Large Language Models (LLMs) extend this capability to semantic quality assessment. LLMs can identify contextual anomalies in relational databases that rule-based methods overlook, such as detecting implausible relationships, temporal inconsistencies, or domain-specific violations [274]. OpenAI’s DaVinci model has demonstrated the capacity to assess temporal, geographic, and technological coverage in datasets, reducing practitioner bias and improving metadata completeness [275]. These capabilities suggest that AI-driven quality assurance may soon rival or exceed human expert judgment in many contexts.

However, adoption remains limited. A systematic review of 151 data quality tools found that most focus on preparing data for AI rather than leveraging AI to improve data quality, highlighting a gap between technological capability and market implementation. Barriers include user acceptance concerns (security, trust, transparency), insufficient organizational skills, and management hesitancy to adopt AI-driven processes [22].

Data Quality for AI: Conversely, AI systems impose stringent quality requirements on their training data. High-quality inputs are crucial for ensuring model accuracy, fairness, robustness, and scalability [276]. Models trained on flawed data produce unreliable predictions, amplify biases, and degrade over time as data distributions shift. Ensuring data quality for AI, therefore, requires:

- Rigorous preprocessing, profiling, and cleansing to eliminate inaccuracies, duplicates, and missing values [277]
- Bias detection and mitigation at the data sourcing stage to prevent algorithmic discrimination [278]
- Continuous monitoring and retraining to address model drift and maintain representativeness [279,280]
- Comprehensive metadata and provenance documentation to support explainability and accountability [281]

The 1-10-100 rule—prevention costs \$1, correction costs \$10, failure costs \$100—applies with particular force in AI contexts, where the cost of failure can include not only financial losses but also reputational damage, legal liability, and social harm [33,101].

AI-based frameworks can support this lifecycle through layered architectures, including data ingestion, quality assessment (e.g., Isolation Forests, BERT for anomaly detection), cleansing and correction (e.g., KNN imputation, LSTM for time series), and continuous monitoring with feedback loops [282–285]. Advanced methods, such as the Denoising Self-Attention Network (DSAN), leverage self-attention and multitask learning to impute mixed-type tabular data, predicting substitutes for both numerical and categorical fields with high accuracy. Data fusion techniques integrate heterogeneous sources, thereby reducing uncertainty and providing more comprehensive representations for model training [285].

Ultimately, sustainable AI governance depends on closing the loop between these two directions. AI tools that improve data quality must themselves be trained on high-quality data, validated for fairness and robustness, and monitored for drift and degradation. This recursive relationship highlights that data quality and AI governance are not separate concerns, but rather intertwined imperatives within the contemporary data ecosystem.

8.5. Synthesis: Toward Integrated AI-Era Data Quality Governance

The AI era demands a fundamental reconceptualization of data quality. Traditional frameworks emphasizing accuracy, completeness, and consistency remain necessary but insufficient. AI-era governance must additionally address algorithmic fairness, model explainability, continuous monitoring, bias mitigation, and ethical accountability. These requirements cannot be retrofitted onto legacy standards; instead, they must be integrated systematically throughout the data and AI lifecycle.

Key principles for AI-era data quality governance include:

1. Lifecycle Integration: Quality assurance must be embedded at every stage—from data sourcing and preparation through model development, deployment, and continuous monitoring—rather than treated as a pre-analytical step [286–291]
2. Ethical Embeddedness: Fairness, transparency, and accountability are not external compliance requirements but intrinsic quality dimensions that must be assessed, measured, and continuously validated [292–301].
3. Continuous Adaptation: AI systems operate in dynamic environments where data distributions shift and new risks emerge. Governance frameworks must incorporate feedback loops, drift detection, and iterative policy updates to remain effective [302–309].
4. Sociotechnical Approach: Technology alone—even sophisticated AI-driven validation systems—cannot resolve governance challenges. Sustainable quality requires integration of technical solutions with organizational design, clear accountability structures, adequate training, and committed leadership [309–311].
5. Bidirectional Innovation: Organizations must leverage AI to improve data quality at scale while simultaneously ensuring that AI systems themselves are trained, validated, and monitored using high-quality, ethically sourced data [305,310,312,313].

The convergence of data quality management and AI ethics represents one of the defining challenges and opportunities of the contemporary data landscape. Organizations that successfully navigate this paradigm shift—by embedding ethical considerations into technical quality frameworks, leveraging AI for continuous assurance, and maintaining rigorous governance throughout the AI lifecycle—will achieve competitive advantages, mitigate risks, and contribute to the development of trustworthy, socially beneficial AI systems. Conversely, those that persist with outdated, compliance-focused approaches risk algorithmic failures, regulatory penalties, reputational damage, and social harm—consequences far exceeding the investments required for comprehensive, forward-looking.

9. Discussion

9.1. *Synthesis of Literature: What Is Agreed upon, What Is Debated*

This review confirms the persistence of inconsistencies in defining and applying data quality dimensions across various sectors. Miller’s hierarchical model offers a promising path by harmonizing terminology and adding new dimensions relevant to modern contexts such as IoT and machine learning [62]. At the same time, AI and machine learning are transforming governance by automating profiling, anomaly detection, and metadata management, reducing manual intervention and improving compliance. However, a systematic review of 151 tools found that most focus on “data quality for AI,” while few exploit “AI for data quality” [108], highlighting a gap between current market offerings and the need for intelligent, automated DQM.

The FAIR principles—Findable, Accessible, Interoperable, and Reusable—provide a widely recognized foundation for transparent management [237]. Initially developed for research data, these principles ensure that datasets are discoverable through rich metadata, accessible with clear usage conditions, interoperable through standardized formats, and reusable across different contexts. Applied not only to data but also to software, they ensure reproducibility by making workflows, tools, and dependencies openly accessible [26,314]. FAIR should be viewed as a flexible, community-driven framework for creating “AI-ready” digital objects that machines can process with minimal human intervention [238], underscoring that alignment with FAIR principles requires sustained organizational commitment beyond technical implementation. [239] The broader reproducibility crisis has eroded trust in science [315], prompting calls for open science interventions that emphasize transparency in data practices to strengthen accountability [316]. Together, these findings position data quality as integral to reproducibility, FAIR compliance, and ongoing scientific reform.

9.2. *Emerging Trends and Lessons from Failures*

Semantic-driven technologies, including ontologies and knowledge graphs, emphasize semantics as a standalone dimension of data quality [20]. Ethical principles are increasingly embedded directly in AI systems through privacy-preserving algorithms such as federated learning and homomorphic encryption, which minimize risks by training or computing on distributed, encrypted data [207]. This reflects a shift toward privacy-by-design architectures, making governance an integral part of technical systems.

Large Language Models (LLMs) extend this trend, enabling semantic-level assessments of data quality. Seabra demonstrated that LLMs can identify contextual anomalies in relational databases that are overlooked by rule-based methods [274,316]. Similarly, OpenAI’s DaVinci model has demonstrated the capacity to assess temporal, geographic, and technological coverage, thereby reducing practitioner bias. [317] Multi-agent frameworks further automate the lifecycle, generating and refining PySpark validation scripts with minimal human input [216]. While these advances promise efficiency, they also raise ethical challenges. AI-driven governance must strike a balance between automation and

safeguards, such as fairness and explainability [76,117], as algorithmic decision-making can amplify biases and erode accountability [257,259].

Failures provide further lessons. Poor timeliness, granularity, or completeness consistently impair CRM analytics. For example, a study reported that 64% of customer data had a timeliness score of ≤ 0.3 , producing outdated segmentation and ineffective campaigns [21]. Such cases reinforce that sustainable progress depends on continuous monitoring, systematic validation, and robust governance. Regulatory penalties and reputational damage highlight the cost of neglect [13]. These lessons show that innovation must be coupled with organizational learning to avoid recurring pitfalls.

9.3. Research Gaps

Despite progress, key gaps remain. The organizational dimension of information quality in public administration is underexplored, as governance structures, staff skills, and administrative processes strongly influence reliability, but are mainly insufficiently studied [28]. Promising theoretical tools, such as sextuple information theory and rough set approaches, are still confined to narrow domains, such as judicial documents [31], and require testing in broader contexts (healthcare, publishing, administration).

Measurement standards also lack harmonization. The WHO's Data Quality Review toolkit [318] proposes a facility-level framework, but adoption remains uneven across regions, while sector-specific standards remain fragmented. In healthcare, real-world data often lack standardized definitions, which hinders analytics and research. Qualitative work in Ethiopian public facilities reveals that inadequate training and infrastructure compromise timeliness, completeness, and accuracy [124], underscoring persistent vulnerabilities in the public sector.

Similarly, SMEs remain underexplored. A study of 85 UK SMEs found that deficits in technical skills and resources limit the adoption of robust practices [125]. More comprehensive empirical research is needed to determine whether current frameworks effectively support data-driven decision-making in resource-constrained settings.

9.4. Implications for Practitioners and Policymakers

For practitioners, embedding quality mechanisms directly into business process models offers a sustainable approach to governance. Validation and monitoring should be built into workflows rather than treated as external add-ons [20]. Reusable “information quality fragments” reduce modeling time, provide explicit checkpoints, and increase confidence, while their validation by domain experts demonstrates feasibility [27].

For policymakers and standard setters, these insights underscore the importance of explicitly validating data quality in process modeling guidelines to mitigate systemic risks. Embedding FAIR principles—ensuring data are Findable, Accessible, Interoperable, and Reusable—remains a practical pathway toward reproducibility and accountability in research [237]. Incorporating FAIR principles into governance frameworks not only ensures compliance with regulatory and ethical standards but also enhances the long-term value of datasets by increasing usability [319].

10. Conclusions

This review demonstrates that data quality is a multidimensional concept that extends beyond technical accuracy to include contextual, representational, and accessibility dimensions, as well as emerging attributes such as governance and semantics. Its central contribution has been to connect classic theoretical frameworks with real-world case studies and contemporary challenges in governance, ethics, and artificial intelligence. Ensuring data

quality is therefore not only a technical task but also a strategic and socio-organizational imperative for scientific reproducibility and organizational trust.

The literature consistently demonstrates that poor data quality undermines decision-making, wastes resources, erodes trust, and yields harmful outcomes. Case studies from finance, healthcare, logistics, and aerospace confirm the pervasive and costly consequences of inadequate quality. Effective practices cannot rely solely on post-hoc cleaning; they must embed validation and governance across the entire data lifecycle—from collection to continuous monitoring and ethical oversight. Standards such as ISO 8000 and ISO 25012, as well as newer hierarchical models, provide practical foundations; however, fragmentation in terminology and uneven adoption persist.

Key challenges remain: the proliferation of heterogeneous data sources, the velocity of big data, and entrenched organizational silos. While AI-driven validation tools and governance frameworks offer promising solutions, cost-benefit trade-offs often prevent systematic adoption.

For practitioners, high-quality data should be treated as a strategic asset. For researchers and policymakers, stronger integration with data ethics and greater attention to underexplored contexts, such as healthcare systems and SMEs, are urgently needed.

Finally, aligning practices with FAIR principles reinforces transparency, reproducibility, and long-term societal impact.

Author Contributions: Conceptualization, E.A.-O., F.G.-G. and I.A.-O.; methodology, E.A.-O., F.G.-G. and I.A.-O.; formal analysis, F.G.-G. and I.A.-O.; investigation, M.G.-A. and L.G.-A.; resources, E.A.-O.; data curation, M.G.-A. and L.G.-A.; writing—original draft preparation, M.G.-A. and L.G.-A.; writing—review and editing, E.A.-O., F.G.-G., I.A.-O., M.G.-A. and L.G.-A.; visualization, M.G.-A. and L.G.-A.; supervision, F.G.-G. and I.A.-O.; project administration, F.G.-G. and I.A.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: In the first revision of this manuscript, the authors used Grammarly Professional for grammar, spelling correction, and polishing the text, and ChatGPT version 5 for maintaining the flow of the text. The authors have reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full Term
AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
BI	Business Intelligence
CLV	Customer Lifetime Value
CRM	Customer Relationship Management
DQA	Data Quality Assessment
DQI	Data Quality Index
DQM	Data Quality Management
DSAN	Denoising Self-Attention Network
DSS	Decision Support Systems
ECCMA	Electronic Commerce Code Management Association
FAIR	Findability, Accessibility, Interoperability, and Reusability
FAIR4RS	FAIR for Research Software
GDPR	General Data Protection Regulation

IoT	Internet of Things
ISO	International Organization for Standardization
KNN	K-Nearest Neighbors
LLMs	Large Language Models
LSTM	Long Short-Term Memory
NATO	North Atlantic Treaty Organization
NCS	NATO Codification System
RFM	Recency, Frequency, and Monetary
SaaS	Software as a Service
SMEs	Small and Medium-sized Enterprises
WHO	World Health Organization

References

- Bernardi, F.; Andrade Alves, D.; Crepaldi, N.Y.; Yamada, D.B.; Lima, V.; Costa Rijo, R.P.C.L. Data Quality in Health Research: A Systematic Literature Review. *medRxiv* **2022**, 2022, 22275804.
- Elahi, E. Data Quality in Healthcare—Benefits, Challenges, and Steps for Improvement. Available online: <https://dataladder.com/data-quality-in-healthcare-data-systems/> (accessed on 16 August 2024).
- Ali, S.M.; Naureen, F.; Noor, A.; Kamel Boulous, M.N.; Aamir, J.; Ishaq, M.; Anjum, N.; Ainsworth, J.; Rashid, A.; Majidulla, A.; et al. Data Quality: A Negotiator between Paper-Based and Digital Records in Pakistan’s TB Control Program. *Data* **2018**, *3*, 27. [CrossRef]
- Chen, H.; Hailey, D.; Wang, N.; Yu, P. A Review of Data Quality Assessment Methods for Public Health Information Systems. *Int. J. Environ. Res. Public Health* **2014**, *11*, 5170–5207. [CrossRef]
- ISO 8000-1:2022(en); Data Quality—Part 1: Overview. International Organization for Standardization (ISO): Geneva, Switzerland, 2022. Available online: <https://www.iso.org/obp/ui/#iso:std:iso:8000:-1:ed-1:v1:en> (accessed on 20 August 2025).
- ECCMA. What Is ISO 8000? Available online: <https://eccma.org/what-is-iso-8000/> (accessed on 20 August 2025).
- ISO 8000-1:2011; Data Quality—Part 1: Overview. International Organization for Standardization ISO: Geneva, Switzerland, 2011.
- ISO 8000-110:2009; Data Quality—Part 110: Master Data: Exchange of Characteristic Data: Syntax, Semantic Encoding, and Conformance to Data Specification. International Organization for Standardization ISO: Geneva, Switzerland, 2009.
- ISO/IEC 25012:2008; Software Engineering—Software Product Quality Requirements and Evaluation (SQuaRE)—Data Quality Model. ISO: Geneva, Switzerland, 2008. Available online: <https://www.iso.org/standard/35736.html> (accessed on 27 October 2025).
- Petrović, M. Data Quality in Customer Relationship Management (CRM): Literature Review. *Strateg. Manag.* **2020**, *25*, 40–47. [CrossRef]
- Henderson, D.; Earley, S.; Sykora, E.; Smith, E. *DAMA-DMBOOK Data Management Body of Knowledge*, 2nd ed.; DAMA International: Basking Ridge, NJ, USA, 2017.
- Alshawhi, S.; Missi, F.; Irani, Z. Organisational, Technical and Data Quality Factors in CRM Adoption—SMEs Perspective. *Ind. Mark. Manag.* **2011**, *40*, 376–383. [CrossRef]
- Henderson, D.; Earley, S.; Sykora, E.; Smith, E. (Eds.) Data Quality. In *DAMA-DMBOOK Data Management Body of Knowledge*; DAMA International: Basking Ridge, NJ, USA, 2017; pp. 551–611.
- Strong, D.M.; Lee, Y.W.; Wang, R.Y. Data Quality in Context. *Commun. ACM* **1997**, *40*, 103–110. [CrossRef]
- Wang, R.Y.; Strong, D.M. Beyond Accuracy: What Data Quality Means to Data Consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [CrossRef]
- Benson, P. NATO Codification System as the Foundation for ISO 8000, the International Standard for Data Quality. *Oil IT J.* **2008**, *1*, 1–4.
- Pipino, L.L.; Lee, Y.W.; Wang, R.Y. Data Quality Assessment. *Commun. ACM* **2002**, *45*, 211–218. [CrossRef]
- Ehrlinger, L.; Wöß, W. A Survey of Data Quality Measurement and Monitoring Tools. *Front. Big Data* **2022**, *5*, 850611. [CrossRef]
- Haug, A.; Zachariassen, F.; van Liempd, D. The Costs of Poor Data Quality. *J. Ind. Eng. Manag.* **2011**, *4*, 168–193. [CrossRef]
- Vaknin, M.; Filipowska, A. *Information Quality Framework for the Design and Validation of Data Flow Within Business Processes-Position Paper*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 158–168. [CrossRef]
- Suh, Y. Exploring the Impact of Data Quality on Business Performance in CRM Systems for Home Appliance Business. *IEEE Access* **2023**, *11*, 116076–116089. [CrossRef]
- Tamm, H.C.; Nikiforova, A. From Data Quality for AI to AI for Data Quality: A Systematic Review of Tools for AI-Augmented Data Quality Management in Data Warehouses. *arXiv* **2025**, arXiv:2406.10940.

23. Bernardo, B.M.V.; Mamede, H.S.; Barroso, J.M.P.; dos Santos, V.M.P.D. Data Governance & Quality Management—Innovation and Breakthroughs across Different Fields. *J. Innov. Knowl.* **2024**, *9*, 100598. [\[CrossRef\]](#)
24. Nguyen, T.; Nguyen, H.-T.; Nguyen-Hoang, T.-A. Data Quality Management in Big Data: Strategies, Tools, and Educational Implications. *J. Parallel Distrib. Comput.* **2025**, *200*, 105067. [\[CrossRef\]](#)
25. Nicholson, N.; Negrao Carvalho, R.; Štotl, I. A FAIR Perspective on Data Quality Frameworks. *Data* **2025**, *10*, 136. [\[CrossRef\]](#)
26. Lamprecht, A.-L.; Garcia, L.; Kuzak, M.; Martinez, C.; Arcila, R.; Martin Del Pico, E.; Dominguez Del Angel, V.; van de Sandt, S.; Ison, J.; Martinez, P.A.; et al. Towards FAIR Principles for Research Software. *Data Sci.* **2020**, *3*, 37–59. [\[CrossRef\]](#)
27. Lopes, C.S.; Silveira, D.S.D.; Araujo, J. Business Processes Fragments to Promote Information Quality. *Int. J. Qual. Reliab. Manag.* **2021**, *38*, 1880–1901. [\[CrossRef\]](#)
28. Oliychenko, I.; Ditkovska, M. Improving Information Quality in E-Government of Ukraine. *Electron. Gov. Int. J.* **2023**, *19*, 146. [\[CrossRef\]](#)
29. Xu, J.; Tang, J.; Ma, X.; Xu, B.; Shen, Y.; Qiao, Y. Objective Information Theory: A Sextuple Model and 9 Kinds of Metrics. *Comput. Sci. Math.* **2014**, *2014*, 793–802.
30. Lian, H.; He, T.; Qin, Z.; Li, H.; Liu, J. Research on the Information Quality Measurement of Judicial Documents. In Proceedings of the 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), Lisbon, Portugal, 16–20 July 2018; pp. 177–181.
31. Chue Hong, N.P.; Aragon, S.; Hettrick, S.; Jay, C. The Future of Research Software Is the Future of Research. *Patterns* **2025**, *6*, 101322. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Even, A.; Shankaranarayanan, G.; Berger, P.D. Evaluating a Model for Cost-Effective Data Quality Management in a Real-World CRM Setting. *Decis. Support Syst.* **2010**, *50*, 152–163. [\[CrossRef\]](#)
33. Foote, K. The Impact of Poor Data Quality (and How to Fix It). Available online: <https://www.dataversity.net/the-impact-of-poor-data-quality-and-how-to-fix-it/> (accessed on 26 August 2024).
34. Payton, F.C.; Zahay, D. Understanding Why Marketing Does Not Use the Corporate Data Warehouse for CRM Applications. *J. Database Mark. Cust. Strategy Manag.* **2003**, *10*, 315–326. [\[CrossRef\]](#)
35. Bidlack, C.; Wellman, M.P. Exceptional Data Quality Using Intelligent Matching and Retrieval. *AI Mag.* **2010**, *31*, 65–73. [\[CrossRef\]](#)
36. Schäffer, T.; Beckmann, H. *Trendstudie Stammdatenqualität 2013: Erhebung der Aktuellen Situation zur Stammdatenqualität in Unternehmen und Daraus Abgeleitete Trends [Trend Study Master Data Quality 2013: Inquiry of the Current Situation of Master Data Quality in Companies and Derived Trends]*; Steinbeis-Edition: Stuttgart, Germany, 2014.
37. Fisher, C.W.; Lauria, E.J.M.; Matheus, C.C. An Accuracy Metric. *J. Data Inf. Qual.* **2009**, *1*, 1–21. [\[CrossRef\]](#)
38. Kelka, H. Supply Chain Resilience Navigating Disruptions Through Strategic Inventory Management. Bachelor's Thesis, Metropolia University of Applied Sciences, Helsinki, Finland, 2024.
39. Al-Harrasi, A.S.; Adarbah, H.Y.; Al-Badi, A.H.; Shaikh, A.K.; Al-Shihi, H.; Al-Barrak, A. Exploring the Adoption of Big Data Analytics in the Oil and Gas Industry: A Case Study. *J. Bus. Commun. Technol.* **2024**, *3*, 1–16. [\[CrossRef\]](#)
40. Joseph, M.; Kumar, D.P.; Keerthana, J.K. Stock Market Analysis and Portfolio Management. In Proceedings of the 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 14–15 March 2024; pp. 1–6. [\[CrossRef\]](#)
41. Purohit, P.; Al Nuaimi, F.; Nakkolakkal, S. Data Governance, Privacy, Data Sharing Challenges. In Proceedings of the SPE Gas & Oil Technology Showcase and Conference, Dubai, United Arab Emirates, 7–9 May 2024. [\[CrossRef\]](#)
42. UTradeAlgos. The Importance of Real-Time Data in Algo Trading Software. Available online: <https://utradealogs.com/blog/the-importance-of-real-time-data-in-algo-trading-software/> (accessed on 27 August 2024).
43. Antwi, B.O.; Adelakun, B.O.; Eziefule, A.O. Transforming Financial Reporting with AI: Enhancing Accuracy and Timeliness. *Int. J. Adv. Econ.* **2024**, *6*, 205–223. [\[CrossRef\]](#)
44. Nwaimo, C.S.; Adegbola, A.E.; Adegbola, M.D.; Adeusi, K.B. Evaluating the Role of Big Data Analytics in Enhancing Accuracy and Efficiency in Accounting: A Critical Review. *Financ. Account. Res. J.* **2024**, *6*, 877–892. [\[CrossRef\]](#)
45. Judijanto, L.; Edtiyarsih, D.D. The Effect of Company Policy, Legal Compliance, and Information Technology on Audit Report Accuracy in the Textile Industry in Tangerang. *West Sci. Account. Financ.* **2024**, *2*, 287–298. [\[CrossRef\]](#)
46. Ehsani-Moghaddam, B.; Martin, K.; Queenan, J.A. Data Quality in Healthcare: A Report of Practical Experience with the Canadian Primary Care Sentinel Surveillance Network Data. *Health Inf. Manag. J.* **2021**, *50*, 88–92. [\[CrossRef\]](#)
47. Lorence, D. Measuring Disparities in Information Capture Timeliness Across Healthcare Settings: Effects on Data Quality. *J. Med. Syst.* **2003**, *27*, 425–433. [\[CrossRef\]](#)
48. Mashoufi, M.; Ayatollahi, H.; Khorasani-Zavareh, D.; Talebi Azad Boni, T. Data Quality in Health Care: Main Concepts and Assessment Methodologies. *Methods Inf. Med.* **2023**, *62*, 005–018. [\[CrossRef\]](#)
49. Wager, K.A.; Schaffner, M.J.; Foulois, B.; Swanson Kazley, A.; Parker, C.; Walo, H. Comparison of the Quality and Timeliness of Vital Signs Data Using Three Different Data-Entry Devices. *CIN Comput. Inform. Nurs.* **2010**, *28*, 205–212. [\[CrossRef\]](#) [\[PubMed\]](#)

50. Alzghoul, A.; Khaddam, A.A.; Abousweilem, F.; Irtaimah, H.J.; Alshaar, Q. How Business Intelligence Capability Impacts Decision-Making Speed, Comprehensiveness, and Firm Performance. *Inf. Dev.* **2024**, *40*, 220–233. [\[CrossRef\]](#)
51. Kusumawardhani, F.K.; Ratmono, D.; Wibowo, S.T.; Darsono, D.; Widyatmoko, S.; Rokhman, N. The Impact of Digitalization in Accounting Systems on Information Quality, Cost Reduction and Decision Making: Evidence from SMEs. *Int. J. Data Netw. Sci.* **2024**, *8*, 1111–1116. [\[CrossRef\]](#)
52. GOV.UK. *Hidden Costs of Poor Data Quality Tackling Data Quality Saves Money and Reduces Risk*; Government Data Quality Hub: London, UK, 2021.
53. Sattler, K.-U. Data Quality Dimensions. In *Encyclopedia of Database Systems*; Springer: New York, NY, USA, 2016; pp. 1–5. [\[CrossRef\]](#)
54. Enterprise Big Data Framework. Understanding Data Quality: Ensuring Accuracy, Reliability, and Consistency. Available online: <https://www.bigdataframework.org/knowledge/understanding-data-quality/> (accessed on 27 August 2024).
55. Chen, B. What is Data Relevance? Definition, Examples, and Best Practices. Available online: <https://www.metaplane.dev/blog/data-relevance-definition-examples> (accessed on 27 August 2024).
56. IBM. What Is Data Quality? Available online: <https://www.ibm.com/think/topics/data-quality> (accessed on 25 October 2025).
57. Okembo, C.; Morales, J.; Lemmen, C.; Zevenbergen, J.; Kuria, D. A Land Administration Data Exchange and Interoperability Framework for Kenya and Its Significance to the Sustainable Development Goals. *Land* **2024**, *13*, 435. [\[CrossRef\]](#)
58. Bammidi, T.; Gutta, L.; Kotagiri, A.; Samayamantri, L.; Vaddy, R. The Crucial Role of Data Quality in Automated Decision-Making Systems. *Int. J. Manag. Educ. Sustain. Dev.* **2024**, *7*, 1–22.
59. Yandrapalli, V. AI-Powered Data Governance: A Cutting-Edge Method for Ensuring Data Quality for Machine Learning Applications. In Proceedings of the 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Vellore, India, 22–23 February 2024; pp. 1–6. [\[CrossRef\]](#)
60. Van Iddekinge, C.H.; Ployhart, R.E. Developments in the Criterion-related Validation of Selection Procedures: A Critical Review and Recommendations for Practice. *Pers. Psychol.* **2008**, *61*, 871–925. [\[CrossRef\]](#)
61. Redman, T.C. Bad Data Costs the U.S. \$3 Trillion per Year. Available online: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year> (accessed on 14 September 2025).
62. Miller, R.; Whelan, H.; Chrubasik, M.; Whittaker, D.; Duncan, P.; Gregório, J. A Framework for Current and New Data Quality Dimensions: An Overview. *Data* **2024**, *9*, 151. [\[CrossRef\]](#)
63. Gartner, Inc. Data Quality: Why It Matters and How to Achieve It. Available online: <https://www.gartner.com/en/data-analytics/topics/data-quality> (accessed on 20 August 2025).
64. Albrecht, R.; Overbeek, S.; van de Weerd, I. Designing a Data Quality Management Framework for CRM Platform Delivery and Consultancy. *SN Comput. Sci.* **2023**, *4*, 742. [\[CrossRef\]](#)
65. Nilashi, M.; Abumalloh, R.A.; Ahmadi, H.; Samad, S.; Alrizq, M.; Abosag, H.; Alghamdi, A. The Nexus Between Quality of Customer Relationship Management Systems and Customers' Satisfaction: Evidence from Online Customers' Reviews. *Heliyon* **2023**, *9*, e21828. [\[CrossRef\]](#)
66. Nikiforova, A. Open Data Quality Evaluation: A Comparative Analysis of Open Data in Latvia. *Balt. J. Mod. Comput.* **2018**, *6*, 363–386. [\[CrossRef\]](#)
67. Southehal, P. *Data Quality: Empowering Businesses with Analytics and AI*; John Wiley & Sons: Hoboken, NJ, USA, 2023.
68. Cornford, S.L.; Wheeler, A.; Feather, M.S.; Plante, J.F. Assurance Equations: A Cost and Criticality Model for Optimizing Quality Assurance Surveillance. In Proceedings of the 2022 IEEE Aerospace Conference (AERO), Big Sky, MT, USA, 5–12 March 2022; pp. 1–13. [\[CrossRef\]](#)
69. Moore, B. How Bad Data Is Ruining Personalized Customer Experiences—And What to Do About It. Available online: <https://www.infoverity.com/en/blog/how-bad-data-is-ruining-personalized-customer-experiences-and-what-to-do-about-it/> (accessed on 7 November 2025).
70. Theodorakopoulos, L.; Theodoropoulou, A. Leveraging Big Data Analytics for Understanding Consumer Behavior in Digital Marketing: A Systematic Review. *Hum. Behav. Emerg. Technol.* **2024**, *2024*, 3641502. [\[CrossRef\]](#)
71. Alves Gomes, M.; Meisen, T. A Review on Customer Segmentation Methods for Personalized Customer Targeting in E-Commerce Use Cases. *Inf. Syst. E-Bus. Manag.* **2023**, *21*, 527–570. [\[CrossRef\]](#)
72. Fu, Q.; Nicholson, G.L.; Easton, J.M. Understanding Data Quality in a Data-Driven Industry Context: Insights from the Fundamentals. *J. Ind. Inf. Integr.* **2024**, *42*, 100729. [\[CrossRef\]](#)
73. Sun, B. Data-Driven Personalized Marketing Strategy Optimization Based on User Behavior Modeling and Predictive Analytics: Sustainable Market Segmentation and Targeting. *PLoS ONE* **2025**, *20*, e0328151. [\[CrossRef\]](#)
74. The Information Difference Ltd.; Experian. *The Data Quality Landscape—Q1 2023*; The Information Difference Ltd.: York, UK, 2023.
75. Validity. *The State of CRM Data Management in 2024*; Validity: Boston, MA, USA, 2024.
76. Rahm, E.; Do, H. Data Cleaning: Problems and Current Approaches. *IEEE Data Eng. Bull.* **2000**, *23*, 3–14.
77. Nagle, T.; Redman, T.C.; Sammon, D. Only 3% of Companies' Data Meets Basic Quality Standards. *Harv. Bus. Rev.* **2017**, *95*, 2–5.

78. Ahani, A.; Rahim, N.Z.A.; Nilashi, M. Forecasting Social CRM Adoption in SMEs: A Combined SEM-Neural Network Method. *Comput. Hum. Behav.* **2017**, *75*, 560–578. [\[CrossRef\]](#)
79. Delone, W.H.; McLean, E.R. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *J. Manag. Inf. Syst.* **2003**, *19*, 9–30. [\[CrossRef\]](#)
80. Azeroual, O.; Saake, G.; Abuosba, M.; Schöpfel, J. Data Quality as a Critical Success Factor for User Acceptance of Research Information Systems. *Data* **2020**, *5*, 35. [\[CrossRef\]](#)
81. Redman, T.C. To Improve Data Quality, Start at the Source. *Harv. Business Rev.* **2020**. Available online: <https://hbr.org/2020/02/to-improve-data-quality-start-at-the-source> (accessed on 27 July 2025).
82. Gatzert, N. The Impact of Corporate Reputation and Reputation Damaging Events on Financial Performance: Empirical Evidence from the Literature. *Eur. Manag. J.* **2015**, *33*, 485–499. [\[CrossRef\]](#)
83. Peña-García, N.; Losada-Otálora, M.; Auza, D.P.; Cruz, M.P. Reviews, Trust, and Customer Experience in Online Marketplaces: The Case of Mercado Libre Colombia. *Front. Commun.* **2024**, *9*, 1460321. [\[CrossRef\]](#)
84. Rushing, B.; Xu, S.; Fairman, A. From Breach to Bias: Measuring Reputation Value and Trust Recovery after Cyber Incidents in Critical Infrastructure. *Int. J. Crit. Infrastruct. Prot.* **2025**, *50*, 100787. [\[CrossRef\]](#)
85. Açıkgoz, F.Y.; Kayakuş, M.; Zăbavă, B.-Ş.; Kabas, O. Brand Reputation and Trust: The Impact on Customer Satisfaction and Loyalty for the Hewlett-Packard Brand. *Sustainability* **2024**, *16*, 9681. [\[CrossRef\]](#)
86. Nuortimo, K.; Harkonen, J.; Breznik, K. Exploring Corporate Reputation and Crisis Communication. *J. Mark. Anal.* **2024**, *2024*, 1–22. [\[CrossRef\]](#)
87. Nagalakshmi, M.; Sai Sri Charan, Y.; Farooq, B.; Gaur, S.; Saxena, R.; Soni, D. The Role of Brand Image in Strategy. *Adv. Consum. Res.* **2025**, *2*, 623–626.
88. Barakat Ali, M.A. The Effect of Firm's Brand Reputation on Customer Loyalty and Customer Word of Mouth: The Mediating Role of Customer Satisfaction and Customer Trust. *Int. Bus. Res.* **2022**, *15*, 30. [\[CrossRef\]](#)
89. La, S.; Choi, B. The Role of Customer Affection and Trust in Loyalty Rebuilding after Service Failure and Recovery. *Serv. Ind. J.* **2012**, *32*, 105–125. [\[CrossRef\]](#)
90. McCance, L. *Fixing the Foundation: The State of Marketing Data Quality 2025*; Adverity: Vienna, Austria, 2025. Available online: <https://www.adverity.com/state-of-play-research-data-quality-2025> (accessed on 21 July 2025).
91. Zhou, Y.; Shi, J.; Stein, R.; Liu, X.; Baldassano, R.N.; Forrest, C.B.; Chen, Y.; Huang, J. Missing Data Matter: An Empirical Evaluation of the Impacts of Missing EHR Data in Comparative Effectiveness Research. *J. Am. Med. Inf. Assoc* **2023**, *30*, 1246–1256. [\[CrossRef\]](#)
92. Lewis, A.E.; Weiskopf, N.; Abrams, Z.B.; Foraker, R.; Lai, A.M.; Payne, P.R.O.; Gupta, A. Electronic Health Record Data Quality Assessment and Tools: A Systematic Review. *J. Am. Med. Inf. Assoc.* **2023**, *30*, 1730–1740. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Heilbroner, S.P.; Carter, C.; Vidmar, D.M.; Mueller, E.T.; Stumpe, M.C.; Miotto, R. A Self-Supervised Framework for Laboratory Data Imputation in Electronic Health Records. *Commun. Med.* **2025**, *5*, 251. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Kumar, P.; Gupta, V. AI-Driven Market Analysis and Business Intelligence. *Int. J. Res. Manag.* **2024**, *6*, 252–260. [\[CrossRef\]](#)
95. European Securities and Market Authority. *2024 Report on Quality and Use of Data*; European Securities and Market Authority: Paris, France, 2024.
96. Harish, A. When NASA Lost a Spacecraft Due to a Metric Math Mistake. Available online: <https://www.simscale.com/blog/nasa-mars-climate-orbiter-metric/> (accessed on 6 November 2025).
97. Euler, E.A.; Jolly, S.; Curtis, H.H. The Failures of the Mars Climate Orbiter and Mars Polar Lander: A Perspective from the People Involved (Paper AAS 01-074). In Proceedings of the 44th Annual American Astronautical Society Guidance, Navigation, and Control Conference, Harbin, China, 5–7 August 2022; American Astronautical Society: Breckenridge, CO, USA, 2001; pp. 2–22.
98. Abdullah, F. A Case Study on the Mars Climate Orbiter and Mars Polar Lander Failures: What Is the Cost of Underestimating Testing. In *Zenodo*; Zenodo: Geneva, Switzerland, 2025.
99. NASA Tangles with the Metric System. *Science* **1999**, *286*, 2241. [\[CrossRef\]](#)
100. Reichhardt, T. NASA Reworks Its Sums after Mars Fiasco. *Nature* **1999**, *401*, 517. [\[CrossRef\]](#)
101. Davidson, N. The Cost of Poor Data Quality on Business Operations. Available online: <https://lakefs.io/blog/poor-data-quality-business-costs/> (accessed on 26 August 2024).
102. Yackel, R. The Impact of Bad Data: A Case Study on Unity. Available online: <https://www.ibm.com/think/insights/observability-data-benefits> (accessed on 6 November 2025).
103. Xie, J.; Sun, L.; Zhao, Y.F. On the Data Quality and Imbalance in Machine Learning-Based Design and Manufacturing—A Systematic Review. *Engineering* **2025**, *45*, 105–131. [\[CrossRef\]](#)
104. U.S. Government Accountability Office. *Criminal History Records: Additional Actions Could Enhance the Completeness of Records Used for Employment-Related Background Checks*; GAO: Washington, DC, USA, 2015.
105. Lageson, S.; Stewart, R. The Problem with Criminal Records: Discrepancies between State Reports and Private-sector Background Checks. *Criminology* **2024**, *62*, 5–34. [\[CrossRef\]](#)

106. Goggins, B.; DeBacco, D. *Survey of State Criminal History Information Systems, 2020*; Bureau of Justice Statistics: Washington, DC, USA, 2022.
107. Bureau of Justice Statistics. *FY 2023 National Criminal History Improvement Program (NCHIP)*; Bureau of Justice Statistics: Washington, DC, USA, 2023.
108. Wand, Y.; Wang, R. Anchoring Data Quality Dimensions in Ontological Foundations. *Commun. ACM* **1996**, *39*, 86–95. [[CrossRef](#)]
109. LaValle, C.R.; Haas, S.M.; Nolan, J.J. *Testing the Validity of Demonstrated Imputation Methods on Longitudinal NIBRS Data*; West Virginia Criminal Justice Statistical Analysis Center: Charleston, WV, USA, 2014.
110. Prescott, J.J.; Starr, S.B. Expungement of Criminal Convictions: An Empirical Study. *Harv. Law Rev.* **2020**, *133*, 2460–2550. [[CrossRef](#)]
111. Redman, T. *Data Driven: Profiting from Your Most Important Business Asset*; Redman, T., Ed.; Harvard Business Review Press: Cambridge, MA, USA, 2013.
112. Strom, K.J.; Smith, E.L. The Future of Crime Data. *Criminol. Public Policy* **2017**, *16*, 1027–1048. [[CrossRef](#)]
113. Mahendra, P.; Doshi, P.; Verma, A.; Shrivastava, S. A Comprehensive Review of AI and ML in Data Governance and Data Quality. In Proceedings of the 2025 3rd International Conference on Inventive Computing and Informatics (ICICI), Bangalore, India, 4–6 June 2025; pp. 1–6. [[CrossRef](#)]
114. Inmon, W.H. *Building the Data Warehouse*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2005.
115. KPMG. *Managing the Data Challenge in Banking*; KPMG: London, UK, 2014.
116. Jeleel-Ojuade, A. The Role of Information Silos: An Analysis of How the Categorization of Information Creates Silos within Financial Institutions, Hindering Effective Communication and Collaboration. *SSRN Electron. J.* **2024**, *2014*, 4881342. [[CrossRef](#)]
117. European Central Bank (SSM). *Guide on Effective Risk Data Aggregation and Risk Reporting*; European Central Bank (SSM): Frankfurt, Germany, 2024.
118. Basel Committee on Banking Supervision. *Principles for Effective Risk Data Aggregation and Risk Reporting (BCBS 239)*; Bank for International Settlements: Basel, Switzerland, 2013.
119. Dehghani, Z. *Data Mesh: Delivering Data-Driven Value at Scale*; O'Reilly Media: Sebastopol, CA, USA, 2022.
120. Capirossi, J.; Rabier, P. *An Enterprise Architecture and Data Quality Framework*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 67–79. [[CrossRef](#)]
121. Taleb, I.; Serhani, M.A.; Bouhaddioui, C.; Dssouli, R. Big Data Quality Framework: A Holistic Approach to Continuous Quality Management. *J. Big Data* **2021**, *8*, 76. [[CrossRef](#)]
122. Alaqla, M.F. The Impact of IT Governance and Administrative Information Quality on Decision-Making in the Banking Sector. *Corp. Gov. Organ. Behav. Rev.* **2023**, *7*, 171–182. [[CrossRef](#)]
123. Weill, P.; Ross, J. *IT Governance: How Top Performers Manage IT Decision Rights for Superior Results*; Harvard Business School Press: Cambridge, MA, USA, 2004.
124. Khatri, V.; Brown, C.V. Designing Data Governance. *Commun. ACM* **2010**, *53*, 148–152. [[CrossRef](#)]
125. Storey, V.C.; Dewan, R.M.; Freimer, M. Data Quality: Setting Organizational Policies. *Decis. Support Syst.* **2012**, *54*, 434–442. [[CrossRef](#)]
126. Yang, Y. Applications and Challenges of Big Data in Market Analytics. *Trans. Econ. Bus. Manag. Res.* **2024**, *9*, 450–458. [[CrossRef](#)]
127. Gomez-Urbe, C.A.; Hunt, N. The Netflix Recommender System. *ACM Trans. Manag. Inf. Syst.* **2016**, *6*, 1–19. [[CrossRef](#)]
128. Pajkovic, N. Algorithms and Taste-Making: Exposing the Netflix Recommender System's Operational Logics. *Converg. Int. J. Res. New Media Technol.* **2022**, *28*, 214–235. [[CrossRef](#)]
129. Gerber, C. A Consumer Perspective on Netflix's Recommender System. A Qualitative Analysis. Master's Thesis, Erasmus University Rotterdam, Rotterdam, The Netherlands, 2021.
130. Dutta, A. Personalized Content Recommendation Impact on User Engagement of Netflix. *Int. J. Res. Publ. Rev.* **2025**, *6*, 10889–10892.
131. Kahn, M.G.; Raebel, M.A.; Glanz, J.M.; Riedlinger, K.; Steiner, J.F. A Pragmatic Framework for Single-Site and Multisite Data Quality Assessment in Electronic Health Record-Based Clinical Research. *Med Care* **2012**, *50*, S21–S29. [[CrossRef](#)]
132. Daniel, C.; Kalra, D. Clinical Research Informatics: Contributions from 2018. *Yearb Med. Inf.* **2019**, *28*, 203–205. [[CrossRef](#)] [[PubMed](#)]
133. Qualls, L.G.; Phillips, T.A.; Hammill, B.G.; Topping, J.; Louzao, D.M.; Brown, J.S.; Curtis, L.H.; Marsolo, K. Evaluating Foundational Data Quality in the National Patient-Centered Clinical Research Network (PCORnet®). *eGEMS (Gener. Evid. Methods Improv. Patient Outcomes)* **2018**, *6*, 3. [[CrossRef](#)] [[PubMed](#)]
134. Daniel Boie, S.; Meyer-Eschenbach, F.; Schreiber, F.; Giesa, N.; Barrenetxea, J.; Guinemer, C.; Haufe, S.; Krämer, M.; Brunecker, P.; Prasser, F.; et al. A Scalable Approach for Critical Care Data Extraction and Analysis in an Academic Medical Center. *Int. J. Med. Inf.* **2024**, *192*, 105611. [[CrossRef](#)]
135. Ozonze, O.; Scott, P.J.; Hopgood, A.A. Automating Electronic Health Record Data Quality Assessment. *J. Med. Syst.* **2023**, *47*, 23. [[CrossRef](#)] [[PubMed](#)]

136. WHO. *Data Quality Assurance (DQA) Toolkit*; WHO: Geneva, Switzerland, 2022.
137. European Medicines Agency. *Committee for Medicinal Products for Human Use (CHMP). Data Quality Framework for EU Medicines Regulation: 4 Application to Real-World Data*; European Medicines Agency: Amsterdam, The Netherlands, 2024.
138. Hibbert, P.D.; Stewart, S.; Wiles, L.K.; Braithwaite, J.; Runciman, W.B.; Thomas, M.J.W. Improving Patient Safety Governance and Systems through Learning from Successes and Failures: Qualitative Surveys and Interviews with International Experts. *Int. J. Qual. Health Care* **2023**, *35*. [\[CrossRef\]](#)
139. Oktaviana, R.S.; Handayani, P.W.; Hidayanto, A.N.; Siswanto, B.B. Healthcare Data Governance Assessment Based on Hospital Management Perspectives. *Int. J. Inf. Manag. Data Insights* **2025**, *5*, 100342. [\[CrossRef\]](#)
140. Lighterness, A.; Adcock, M.; Scanlon, L.A.; Price, G. Data Quality–Driven Improvement in Health Care: Systematic Literature Review. *J. Med. Internet Res.* **2024**, *26*, e57615. [\[CrossRef\]](#)
141. Payne, A.; Frow, P. Relationship Marketing: Looking Backwards towards the Future. *J. Serv. Mark.* **2017**, *31*, 11–15. [\[CrossRef\]](#)
142. Choudhury, M.M.; Harrigan, P. CRM to Social CRM: The Integration of New Technologies into Customer Relationship Management. *J. Strateg. Mark.* **2014**, *22*, 149–176. [\[CrossRef\]](#)
143. Becker, J.U.; Greve, G.; Albers, S. The Impact of Technological and Organizational Implementation of CRM on Customer Acquisition, Maintenance, and Retention. *Int. J. Res. Mark.* **2009**, *26*, 207–215. [\[CrossRef\]](#)
144. Shum, P.; Bove, L.; Auh, S. Employees’ Affective Commitment to Change. *Eur. J. Mark.* **2008**, *42*, 1346–1371. [\[CrossRef\]](#)
145. Adane, A.; Adege, T.M.; Ahmed, M.M.; Anteneh, H.A.; Ayalew, E.S.; Berhanu, D.; Berhanu, N.; Getnet, M.; Bishaw, T.; Busza, J.; et al. Exploring Data Quality and Use of the Routine Health Information System in Ethiopia: A Mixed-Methods Study. *BMJ Open* **2021**, *11*, e050356. [\[CrossRef\]](#) [\[PubMed\]](#)
146. Tilahun, H.; Abate, B.; Belay, H.; Gebeyehu, A.; Ahmed, M.; Simanesew, A.; Ayele, W.; Mohammedsanni, A.; Knittel, B.; Wondarad, Y. Drivers and Barriers to Improved Data Quality and Data-Use Practices: An Interpretative Qualitative Study in Addis Ababa, Ethiopia. *Glob. Health Sci. Pract.* **2022**, *10* (Suppl. 1), e2100689. [\[CrossRef\]](#)
147. Tolera, A.; Firdisa, D.; Roba, H.S.; Motuma, A.; Kitesa, M.; Abaerei, A.A. Barriers to Healthcare Data Quality and Recommendations in Public Health Facilities in Dire Dawa City Administration, Eastern Ethiopia: A Qualitative Study. *Front. Digit. Health* **2024**, *6*, 1261031. [\[CrossRef\]](#)
148. Gazi, M.A.I.; Al Mamun, A.; Al Masud, A.; Senathirajah, A.R.B.S.; Rahman, T. The Relationship between CRM, Knowledge Management, Organization Commitment, Customer Profitability and Customer Loyalty in Telecommunication Industry: The Mediating Role of Customer Satisfaction and the Moderating Role of Brand Image. *J. Open Innov. Technol. Mark. Complex.* **2024**, *10*, 100227. [\[CrossRef\]](#)
149. Lee, Y.-C.; Wang, Y.-C.; Lu, S.-C.; Hsieh, Y.-F.; Chien, C.-H.; Tsai, S.-B.; Dong, W. An Empirical Research on Customer Satisfaction Study: A Consideration of Different Levels of Performance. *Springerplus* **2016**, *5*, 1577. [\[CrossRef\]](#)
150. Guerola-Navarro, V.; Oltra-Badenes, R.; Gil-Gomez, H.; Gil-Gomez, J.A. Research Model for Measuring the Impact of Customer Relationship Management (CRM) on Performance Indicators. *Econ. Res.-Ekon. Istraživanja* **2021**, *34*, 2669–2691. [\[CrossRef\]](#)
151. Eklof, J.; Podkorytova, O.; Malova, A. Linking Customer Satisfaction with Financial Performance: An Empirical Study of Scandinavian Banks. *Total Qual. Manag. Bus. Excell.* **2020**, *31*, 1684–1702. [\[CrossRef\]](#)
152. Prasad, A. Impact of Poor Data Quality on Business Performance: Challenges, Costs, and Solutions. *SSRN Electron. J.* **2024**. Available online: <https://ssrn.com/abstract=4843991> (accessed on 29 July 2025). [\[CrossRef\]](#)
153. Haverila, M.; Haverila, K.C.; Mohiuddin, M.; Su, Z. The Impact of Quality of Big Data Marketing Analytics (BDMA) on the Market and Financial Performance. *J. Glob. Inf. Manag.* **2022**, *30*, 1–21. [\[CrossRef\]](#)
154. Redyuk, S.; Kaoudi, Z.; Markl, V.; Schelter, S. Automating Data Quality Validation for Dynamic Data Ingestion. In Proceedings of the 24th International Conference on Extending Database Technology, EDBT’21, Nicosia, Cyprus, 23–26 March 2021; pp. 61–72.
155. Syed, R.; Eden, R.; Makasi, T.; Chukwudi, I.; Mamudu, A.; Kamalpour, M.; Kapugama Geeganage, D.; Sadeghianasl, S.; Leemans, S.J.J.; Goel, K.; et al. Digital Health Data Quality Issues: Systematic Review. *J. Med. Internet Res.* **2023**, *25*, e42615. [\[CrossRef\]](#)
156. Barchard, K.A.; Freeman, A.J.; Ochoa, E.; Stephens, A.K. Comparing the Accuracy and Speed of Four Data-Checking Methods. *Behav. Res. Methods* **2020**, *52*, 97–115. [\[CrossRef\]](#)
157. Perez-Castillo, R.; Carretero, A.G.; Caballero, I.; Rodriguez, M.; Piattini, M.; Mate, A.; Kim, S.; Lee, D. DAQUA-MASS: An ISO 8000-61 Based Data Quality Management Methodology for Sensor Data. *Sensors* **2018**, *18*, 3105. [\[CrossRef\]](#)
158. Silva, M.D.S.T.; Correia, S.É.N.; de A. Machado, P.; de Oliveira, V.M. Adoption of Information Technology in Public Administration: A Focus on the Organizational Factors of a Brazilian Federal University. *Teor. Prática Adm.* **2020**, *10*, 138–153. [\[CrossRef\]](#)
159. Yukhno, A. Digital Transformation: Exploring Big Data Governance in Public Administration. *Public Organ. Rev.* **2024**, *24*, 335–349. [\[CrossRef\]](#)
160. Cerrillo-Martínez, A.; Casadesús-de-Mingo, A. Data Governance for Public Transparency. *El Prof. Inf.* **2021**, *30*, e300402. [\[CrossRef\]](#)
161. Lutsenko, K. Digitalisation of Public Administration: Challenges and Prospects. *Health Leadersh. Qual. Life* **2024**, *3*, 434. [\[CrossRef\]](#)
162. OECD. *Developing Skills for Digital Government: A Review of Good Practices Across OECD Governments*; OECD: Paris, France, 2024.

163. Tawil, A.-R.; Mohamed, M.; Schmoor, X.; Vlachos, K.; Haidar, D. Trends and Challenges Towards an Effective Data-Driven Decision Making in UK SMEs: Case Studies and Lessons Learnt from the Analysis of 85 SMEs. *arXiv* **2023**, arXiv:2305.15454. [CrossRef]
164. Mohamed, M.; Weber, P. Trends of Digitalization and Adoption of Big Data & Analytics among UK SMEs: Analysis and Lessons Drawn from a Case Study of 53 SMEs. In Proceedings of the 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Cardiff, UK, 15–17 June 2020; pp. 1–6. [CrossRef]
165. Gates, S. 5 Examples of Bad Data Quality in Business—And How to Avoid Them. Available online: <https://www.montecarlodata.com/blog-bad-data-quality-examples/> (accessed on 26 August 2024).
166. Federal Trade Commission. *Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003*; Federal Trade Commission: Washington, DC, USA, 2015.
167. Schroeder, P. US Consumer Bureau Fines Equifax \$15 Million over Handling of Consumer Disputes. *Reuters* **2025**. Available online: <https://www.reuters.com/business/finance/us-consumer-bureau-fines-equifax-15-million-issues-fixing-consumer-disputes-2025-01-17/> (accessed on 26 August 2024).
168. Mars Climate Orbiter Mishap Investigation Board. *Mars Climate Orbiter Mishap Investigation Board Phase I Report*; Mars Climate Orbiter Mishap Investigation Board: Washington, DC, USA, 2019.
169. Data Ladder. Data Ladder Whitepapers | How Legacy Systems and Bad Data Quality Hinders a Digital Transformation Plan-Data Ladder. Available online: https://dataladder.com/whitepapers/how-legacy-systems-and-bad-data-quality-hinders-a-digital-transformation-plan/?imz_s=9nekd6omo7qd26qrf5hdkithi6%2F (accessed on 9 November 2025).
170. Deepak Veeravalli, S. Legacy System Modernization: Guidelines for Migrating from Legacy Systems to Salesforce: Address Challenges and Implementing Best Practices with Reusable Integration Blueprints. *Int. J. Comput. Sci. Inf. Technol. Res.* **2022**, *3*, 133–144. [CrossRef]
171. Hüner, M.K.; Schierner, A.; Otto, B.; Österle, H. Product Data Quality in Supply Chains: The Case of Beiersdorf. *Electron. Mark.* **2011**, *21*, 141–154. [CrossRef]
172. Rigo, G.-E.; Pedron, C.D.; Caldeira, M.; De Araújo, C.C.S. CRM Adoption in a Higher Education Institution. *J. Inf. Syst. Technol. Manag.* **2016**, *13*, 45–60. [CrossRef]
173. Weerts, D.J.; Ronca, J.M. Characteristics of Alumni Donors Who Volunteer at Their Alma Mater. *Res. High. Educ.* **2008**, *49*, 274–292. [CrossRef]
174. Research Group of the Office of the Privacy Commissioner of Canada. *The Age of Predictive Analytics: From Patterns to Predictions—Office of the Privacy Commissioner of Canada*; Research Group of the Office of the Privacy Commissioner of Canada: Gatineau, QC, Canada, 2012.
175. Biemer, P.P. Data Quality and Inference Errors. In *Big Data and Social Science Data Science Methods and Tools for Research and Practice*; Foster, I., Ghani, R., Jarmin, R., Kreuter, F., Lane, J., Eds.; CRC: Boca Raton, FL, USA, 2020.
176. Butler, D. When Google Got Flu Wrong. *Nature* **2013**, *494*, 155–156. [CrossRef]
177. Lazer, D.; Kennedy, R.; King, G.; Vespignani, A. The Parable of Google Flu: Traps in Big Data Analysis. *Science (1979)* **2014**, *343*, 1203–1205. [CrossRef]
178. Lazer, D.; Kennedy, R.; King, G.; Vespignani, A. Google Flu Trends Still Appears Sick: An Evaluation of the 2013–2014 Flu Season. *SSRN Electron. J.* **2014**, *2014*, 2408560. [CrossRef]
179. Algemene Rekenkamer. *Datagedreven Selectie van Aangiften Door de Belastingdienst | Rapport | Algemene Rekenkamer [Data-Driven Selection of Tax Returns by the Dutch Tax and Customs Administration | Report | Netherlands Court of Audit]*; Algemene Rekenkamer: The Hague, The Netherlands, 2019.
180. OECD. *Tax Administration 3.0: The Digital Transformation of Tax Administration*; OECD: Paris, France, 2020. [CrossRef]
181. Aslett, J. *Tax Administration*; International Monetary Fund: Washington, DC, USA, 2024; Volume 2024. [CrossRef]
182. WiredGov. The Damaging Impact of Poor Quality Data in the Public Sector | Official Press Release. Available online: <https://www.wired-gov.net/wg/content.nsf/industrynews/The+damaging+impact+of+poor+quality+data+in+the+public+sector?open&id=BDEX-6ZFKSP> (accessed on 9 November 2025).
183. Marzullo, A.; Savevski, V.; Menini, M.; Schilirò, A.; Franchellucci, G.; Dal Buono, A.; Bezzio, C.; Gabbiadini, R.; Hassan, C.; Repici, A.; et al. Collecting and Analyzing IBD Clinical Data for Machine-Learning: Insights from an Italian Cohort. *Data* **2025**, *10*, 100. [CrossRef]
184. Tlouyamma, J.; Mokwena, S. Automated Data Quality Control System in Health and Demographic Surveillance System. *Sci. Eng. Technol.* **2024**, *4*, 82–91. [CrossRef]
185. Razzaghi, H.; Goodwin Davies, A.; Boss, S.; Bunnell, H.T.; Chen, Y.; Chrischilles, E.A.; Dickinson, K.; Hanauer, D.; Huang, Y.; Ilunga, K.T.S.; et al. Systematic Data Quality Assessment of Electronic Health Record Data to Evaluate Study-Specific Fitness: Report from the PRESERVE Research Study. *PLoS Digit. Health* **2024**, *3*, e0000527. [CrossRef] [PubMed]
186. Wang, Y.; Hulstijn, J.; Tan, Y.-H. Data Quality Assurance in International Supply Chains: An Application of the Value Cycle Approach to Customs Reporting. *Int. J. Adv. Logist.* **2016**, *5*, 76–85. [CrossRef]

187. Zovko, L. Digitalization in Health Systems in the European Union. Bachelor's Thesis, University of Zagreb, Zagreb, Croatia, 2025.
188. Gelashvili-Luik, T.; Vihma, P.; Pappel, I. Navigating the AI Revolution: Challenges and Opportunities for Integrating Emerging Technologies into Knowledge Management Systems. Systematic Literature Review. *Front. Artif. Intell.* **2025**, *8*, 1595930. [[CrossRef](#)] [[PubMed](#)]
189. Masod, M.Y.B.; Zakaria, S.F. Artificial Intelligence Adoption in the Manufacturing Sector: Challenges and Strategic Framework. *Int. J. Res. Innov. Soc. Sci.* **2024**, *8*, 150–158. [[CrossRef](#)]
190. Kapiki, S.; Pappa, A. Enhancing Healthcare Efficiency: Leveraging Advanced Maintenance Management for Optimal Staff Performance. *J. Health Organ. Manag.* **2025**, *39*, 398–418. [[CrossRef](#)]
191. Davidson, P.L.; Hunt, J.; La Manna, A.; Luke, D.A. Editorial: Impact Evaluation Using the Translational Science Benefits Model Framework in the National Center for Advancing Translational Science Clinical and Translational Science Award Program. *Front. Public Health* **2025**, *13*, 1707595. [[CrossRef](#)] [[PubMed](#)]
192. Ebenso, B.; Namisango, E.; Abejirinde, I.-O.; Allsop, M.J. Editorial: The Scale-up and Sustainability of Digital Health Interventions in Low- and Middle-Income Settings. *Front. Digit. Health* **2025**, *7*, 1634223. [[CrossRef](#)]
193. Shi, Y.; Li, J.; Kou, G.; Tien, J.M.; Berg, D. Merging Artificial Intelligence and Business Applications: Preface for ITQM 2025. *Procedia Comput. Sci.* **2025**, *266*, 1–8. [[CrossRef](#)]
194. Pykes, K. 10 Signs of Bad Data: How to Spot Poor Quality Data. Available online: <https://www.datacamp.com/blog/10-signs-bad-data-quality> (accessed on 26 August 2024).
195. Fu, A.; Shen, T.; Roberts, S.B.; Liu, W.; Vaidyanathan, S.; Marchena-Romero, K.-J.; Lam, Y.Y.P.; Shah, K.; Mak, D.Y.F.; Chin, S.; et al. Optimizing the Efficiency and Effectiveness of Data Quality Assurance in a Multicenter Clinical Dataset. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 835–844. [[CrossRef](#)]
196. Haverila, M.J.; Haverila, K.C. The Influence of Quality of Big Data Marketing Analytics on Marketing Capabilities: The Impact of Perceived Market Performance! *Mark. Intell. Plan.* **2024**, *42*, 346–372. [[CrossRef](#)]
197. Lee, D.-H.; Kim, H. A Self-Attention-Based Imputation Technique for Enhancing Tabular Data Quality. *Data* **2023**, *8*, 102. [[CrossRef](#)]
198. Becerra, M.A.; Tobón, C.; Castro-Ospina, A.E.; Peluffo-Ordóñez, D.H. Information Quality Assessment for Data Fusion Systems. *Data* **2021**, *6*, 60. [[CrossRef](#)]
199. MacDonald, L. Measuring Data Quality: Key Metrics, Processes, and Best Practices. Available online: <https://www.montecarlodata.com/blog-measuring-data-quality-key-metrics-processes-and-best-practices/> (accessed on 27 August 2024).
200. Karkošková, S. Data Governance Model to Enhance Data Quality in Financial Institutions. *Inf. Syst. Manag.* **2023**, *40*, 90–110. [[CrossRef](#)]
201. Sluzki, N. 8 Data Quality Monitoring Techniques & Metrics to Watch. Available online: <https://www.ibm.com/think/topics/data-quality-monitoring-techniques> (accessed on 27 August 2024).
202. Verma, P.; Kumar, V.; Mittal, A.; Rathore, B.; Jha, A.; Rahman, M.S. The Role of 3S in Big Data Quality: A Perspective on Operational Performance Indicators Using an Integrated Approach. *TQM J.* **2023**, *35*, 153–182. [[CrossRef](#)]
203. Woods, C.; Selway, M.; Bikaun, T.; Stumptner, M.; Hodkiewicz, M. An Ontology for Maintenance Activities and Its Application to Data Quality. *Semant. Web* **2024**, *15*, 319–352. [[CrossRef](#)]
204. Stepanenko, R. Data Stewardship Explained: The Backbone of Data Management. Available online: <https://recordlinker.com/data-stewardship-explained/> (accessed on 14 September 2025).
205. Jatin, B. Data Governance for Quality: Policies Ensuring Reliable Data. Available online: <https://www.decube.io/post/data-quality-data-governance> (accessed on 27 August 2024).
206. Hanna, M.G.; Pantanowitz, L.; Jackson, B.; Palmer, O.; Visweswaran, S.; Pantanowitz, J.; Deebajah, M.; Rashidi, H.H. Ethical and Bias Considerations in Artificial Intelligence/Machine Learning. *Mod. Pathol.* **2025**, *38*, 100686. [[CrossRef](#)]
207. Duggireddy, G.B.R. Integrated Data and AI Governance Framework: A Lifecycle Approach to Responsible AI Implementation. *J. Comput. Sci. Technol. Stud.* **2025**, *7*, 771–777.
208. Papagiannidis, E.; Mikalef, P.; Conboy, K. Responsible Artificial Intelligence Governance: A Review and Research Framework. *J. Strateg. Inf. Syst.* **2025**, *34*, 101885. [[CrossRef](#)]
209. Floridi, L.; Taddeo, M. What Is Data Ethics? *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20160360. [[CrossRef](#)]
210. Pahune, S.; Akhtar, Z.; Mandapati, V.; Siddique, K. The Importance of AI Data Governance in Large Language Models. *Big Data Cogn. Comput.* **2025**, *9*, 147. [[CrossRef](#)]
211. Forrest, S. Study Examines Accuracy of Arrest Data in FBI's NIBRS Crime Database. Available online: <https://phys.org/news/2022-02-accuracy-fbi-nibrs-crime-database.html> (accessed on 15 August 2024).
212. Labarrère, N.; Costa, L.; Lima, R.M. Data Science Project Barriers—A Systematic Review. *Data* **2025**, *10*, 132. [[CrossRef](#)]
213. Illinois Criminal Justice Information Authority. *Annual Audit Report for 1982–1983: Data Quality of Computerized Criminal Histories*; National Criminal Justice Reference Service (NCJRS): Rockville, MD, USA, 1983.

214. Bosse, R.C.; Jino, M.; de Franco Rosa, F. *A Study on Data Quality and Analysis in Business Intelligence*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 249–253. [\[CrossRef\]](#)
215. Sienkiewicz, M. From Data Silos to Data Mesh: A Case Study in Financial Data Architecture. In Proceedings of the 36th International Conference, DEXA 2025, Bangkok, Thailand, 25–27 August 2025; pp. 3–20. [\[CrossRef\]](#)
216. Senguttuvan, K.R. Multi-Agent Based Automated Data Quality Engineering. Master's Thesis, Fordham University, New York, NY, USA, 2025.
217. Stamkou, C.; Saprikis, V.; Fragulis, G.F.; Antoniadis, I. User Experience and Perceptions of AI-Generated E-Commerce Content: A Survey-Based Evaluation of Functionality, Aesthetics, and Security. *Data* **2025**, *10*, 89. [\[CrossRef\]](#)
218. Vanam, R.R.; Pingili, R.; Myadaboyina, S.G. AI-Based Data Quality Assurance for Business Intelligence and Decision Support Systems. *Int. J. Emerg. Trends Comput. Sci. Inf. Technol.* **2025**, *6*, 21–29. [\[CrossRef\]](#)
219. Elouataoui, W.; El Mendili, S.; Gahi, Y. An Automated Big Data Quality Anomaly Correction Framework Using Predictive Analysis. *Data* **2023**, *8*, 182. [\[CrossRef\]](#)
220. Pasupuleti, S. AI-Augmented Data Pipelines: Integrating Machine Learning for Intelligent Data Processing. *J. Comput. Sci. Technol. Stud.* **2025**, *7*, 276–283.
221. Dhanekula, A. AI-Driven Business Intelligence Framework for Predictive Decision-Making and Strategic Resource Optimization. *Int. J. Bus. Econ. Insights* **2025**, *05*, 1238–1270. [\[CrossRef\]](#)
222. Tomar, S.; Kadaverugu, R. *Trend Analysis of Long-Term Temperature Data for Prediction of Heat Waves Through Statistical Analysis Using Extreme Value Theory for Climate Disaster Management*; Springer: Singapore, 2025; pp. 91–106. [\[CrossRef\]](#)
223. Cinar, R.F.; Yuksek, G.; Lale, T.; Ekinci, S.; Izci, D.; Ma'arif, A. SHAP-Based Framework for Temporal Detection of Sensor Drift in Gas Sensor Arrays. *J. Robot. Control* **2025**, *6*, 2592–2601.
224. Shafaghat, A. Integrating Artificial Intelligence and Machine Learning to Forecast Air Pollution Impacts on Climate Variability and Public Health. *bioRxiv* **2025**, *2025*, 685968. [\[CrossRef\]](#)
225. Ermilov, A.; Tveritnev, A.; Trusova, A. New Role of Technical Specialists to Enable Digital Transformation in the Petroleum Industry: A Petrophysicist-Based Proof of Concept. In Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, United Arab Emirates, 3–6 November 2025; SPE: Washington, DC, USA, 2025. [\[CrossRef\]](#)
226. ISO 8000-8:2015; Data Quality—Part 8: Information and Data Quality: Concepts and Measuring. ISO: Geneva, Switzerland, 2015.
227. Abhishek, A.; Erickson, L.; Bandopadhyay, T. Data and AI Governance: Promoting Equity, Ethics, and Fairness in Large Language Models. *arXiv* **2025**, arXiv:2508.03970. [\[CrossRef\]](#)
228. Angwin, J.; Larson, J.; Mattu, S.; Kirchner, L.; ProPublica. Machine Bias—ProPublica. Available online: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed on 1 November 2025).
229. Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science* **2019**, *366*, 447–453. [\[CrossRef\]](#)
230. Batool, A.; Zowghi, D.; Bano, M. AI Governance: A Systematic Literature Review. *AI Ethics* **2025**, *5*, 3265–3279. [\[CrossRef\]](#)
231. Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Daumé, H., III; Crawford, K. Datasheets for Datasets. *Commun. ACM* **2021**, *64*, 86–92. [\[CrossRef\]](#)
232. Franklin, G.; Stephens, R.; Piracha, M.; Tiosano, S.; Lehouillier, F.; Koppel, R.; Elkin, P.L. The Sociodemographic Biases in Machine Learning Algorithms: A Biomedical Informatics Perspective. *Life* **2024**, *14*, 652. [\[CrossRef\]](#)
233. Leslie, D. *Understanding Artificial Intelligence Ethics and Safety*; The Alan Turing Institute: London, UK, 2019.
234. Belenguer, L. AI Bias: Exploring Discriminatory Algorithmic Decision-Making Models and the Application of Possible Machine-Centric Solutions Adapted from the Pharmaceutical Industry. *AI Ethics* **2022**, *2*, 771–787. [\[CrossRef\]](#)
235. Cross, T.P.; Wagner, A.; Bibel, D. The Accuracy of Arrest Data in the National Incident-Based Reporting System (NIBRS). *Crime Delinq.* **2023**, *69*, 2484–2507. [\[CrossRef\]](#)
236. Bayram, F.; Ahmed, B.S. Towards Trustworthy Machine Learning in Production: An Overview of the Robustness in MLOps Approach. *ACM Comput. Surv.* **2025**, *57*, 1–35. [\[CrossRef\]](#)
237. Kore, A.; Abbasi Babil, E.; Subasri, V.; Abdalla, M.; Fine, B.; Dolatabadi, E.; Abdalla, M. Empirical data drift detection experiments on real-world medical imaging data. *Nat. Commun.* **2024**, *15*, 1887. [\[CrossRef\]](#) [\[PubMed\]](#)
238. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 16001. [\[CrossRef\]](#)
239. Mons, B.; Schultes, E.; Liu, F.; Jacobsen, A. The FAIR Principles: First Generation Implementation Choices and Challenges. *Data Intell.* **2020**, *2*, 1–9. [\[CrossRef\]](#)
240. Stvilia, B.; Pang, Y.; Lee, D.J.; Gunaydin, F. Data Quality Assurance Practices in Research Data Repositories—A Systematic Literature Review. An Annual Review of Information Science and Technology (ARIST) Paper. *J. Inf. Sci. Technol.* **2025**, *76*, 238–261. [\[CrossRef\]](#)
241. Open Data Institute. *A Framework for AI-Ready Data*; Open Data Institute: London, UK, 2025.

242. Publications Office of the European Union. *Principles and Recommendations to Make Data.Europa.Eu Data More Reusable*; Publications Office of the European Union: Luxembourg, 2022.
243. Clark, T.; Caufield, H.; Parker, J.A.; Al Manir, S.; Amorim, E.; Eddy, J.; Gim, N.; Gow, B.; Goar, W.; Haendel, M.; et al. AI-Readiness for Biomedical Data: Bridge2AI Recommendations. *bioRxiv* **2024**, 2024, 619844. [\[CrossRef\]](#)
244. Hiniduma, K.; Byna, S.; Bez, J.L.; Madduri, R. AI Data Readiness Inspector (AIDRIN) for Quantitative Assessment of Data Readiness for AI. In Proceedings of the 36th International Conference on Scientific and Statistical Database Management, Rennes, France, 10–12 July 2024; ACM: New York, NY, USA, 2024; pp. 1–12. [\[CrossRef\]](#)
245. Ravi, N.; Chaturvedi, P.; Huerta, E.A.; Liu, Z.; Chard, R.; Scourtas, A.; Schmidt, K.J.; Chard, K.; Blaiszik, B.; Foster, I. FAIR Principles for AI Models with a Practical Application for Accelerated High Energy Diffraction Microscopy. *Sci. Data* **2022**, *9*, 657. [\[CrossRef\]](#) [\[PubMed\]](#)
246. ISO/IEC 42001:2023; Information Technology—Artificial Intelligence—Management System. International Organization for Standardization: Geneva, Switzerland, 2023.
247. Morshed, A. Ensuring Trust in Sustainability Financial Reports: The Role of AI and Blockchain in Metadata Standardization. *Manag. Sustain. Arab Rev.* **2025**, 2025, 1–24. [\[CrossRef\]](#)
248. Kamisetty, N.S. Intelligent Cloud-Based KNN Model for Enhancing Data Quality in SAP Financial Systems. *Int. J. Res. Appl. Innov. (IJRAI)* **2025**, *8*, 12909–12914.
249. European Parliament; European Council. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonized Rules on Artificial Intelligence and Amending Certain Legislative Acts (Artificial Intelligence Act)*; European Union: Brussels, Belgium, 2024.
250. Loza Corera, M. Data and Data Governance and Connections to Data Protection Principles in Article 10 of the Artificial Intelligence Act. In *The European Union Artificial Intelligence Act*; CotinoHueso, L., GaLetta, D., Eds.; Editoriale Scientifica: Napoli, Italy, 2025; pp. 595–626.
251. National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST Special Publication 1270; NIST: Gaithersburg, MD, USA, 2023.
252. Rodríguez Valencia, L.; Ochoa Arellano, M.J.; Gutiérrez Figueroa, S.A.; Mur Nuño, C.; Monsalve Piqueras, B.; Corrales Paredes, A.D.V.; Bemposta Rosende, S.; López López, J.M.; Puertas Sanz, E.; Levi Alfaroviz, A. A Systematic Review of Artificial Intelligence Applied to Compliance: Fraud Detection in Cryptocurrency Transactions. *J. Risk Financ. Manag.* **2025**, *18*, 612. [\[CrossRef\]](#)
253. Alotaibi, K.O. Developing a Comprehensive Financial Reporting Governance Framework Using AI Techniques. *Eng. Technol. Appl. Sci. Res.* **2025**, *15*, 29202–29207.
254. Santos, W.D.S.; Coutinho, J.R.; Baião, F.; Miranda Spyrides, G.; Vieira Lopes, H.C. Enhancing Declarative Business Process Management Availability through Generative AI. *Process Sci.* **2025**, *2*, 21. [\[CrossRef\]](#)
255. Ali, S.; Rehman, T.; Saira, S. Exploring Pakistan’s Legal Challenges in Artificial Intelligence Regulation: A Data-Driven Approach. *Crit. Rev. Soc. Sci. Stud.* **2025**, *3*, 1096–1108. [\[CrossRef\]](#)
256. Bayram, F.; Ahmed, B.S.; Hallin, E. Adaptive Data Quality Scoring Operations Framework Using Drift-Aware Mechanism for Industrial Applications. *J. Syst. Softw.* **2024**, *217*, 112184. [\[CrossRef\]](#)
257. Cheong, B.C. Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Front. Hum. Dyn.* **2024**, *6*, 1421273. [\[CrossRef\]](#)
258. Bayram, S.B.; Caliskan, N. Effect of a Game-Based Virtual Reality Phone Application on Tracheostomy Care Education for Nursing Students: A Randomized Controlled Trial. *Nurse Educ. Today* **2019**, *79*, 25–31. [\[CrossRef\]](#)
259. High-Level Expert Group on AI (AI HLEG). *Ethics Guidelines for Trustworthy AI*; European Union: Brussels, Belgium, 2019.
260. Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*; St. Martin’s Press: New York, NY, USA, 2018.
261. Lyu, Q.; Tan, J.; Zapadka, M.E.; Ponnatapura, J.; Niu, C.; Myers, K.J.; Wang, G.; Whitlow, C.T. Translating Radiology Reports into Plain Language Using ChatGPT and GPT-4 with Prompt Learning: Results, Limitations, and Potential. *Vis. Comput. Ind. Biomed. Art* **2023**, *6*, 9. [\[CrossRef\]](#)
262. Kazlaris, I.; Antoniou, E.; Diamantaras, K.; Bratsas, C. From Illusion to Insight: A Taxonomic Survey of Hallucination Mitigation Techniques in LLMs. *AI* **2025**, *6*, 260. [\[CrossRef\]](#)
263. Anh-Hoang, D.; Tran, V.; Nguyen, L.-M. Survey and Analysis of Hallucinations in Large Language Models: Attribution to Prompting Strategies or Model Behavior. *Front. Artif. Intell.* **2025**, *8*, 1622292. [\[CrossRef\]](#) [\[PubMed\]](#)
264. Farquhar, S.; Kossen, J.; Kuhn, L.; Gal, Y. Detecting Hallucinations in Large Language Models Using Semantic Entropy. *Nature* **2024**, *630*, 625–630. [\[CrossRef\]](#)
265. Polyzotis, N.; Zinkevich, M.; Roy, S.; Breck, E.; Whang, S. Data Validation for Machine Learning. In Proceedings of the Second Conference on Machine Learning and Systems, SysML 2019, Stanford, CA, USA, 31 March–2 April 2019.
266. Gautam, A.R. Impact of High Data Quality on LLM Hallucinations. *Int. J. Comput. Appl.* **2025**, *187*, 35–39. [\[CrossRef\]](#)

267. Chelli, M.; Descamps, J.; Lavoué, V.; Trojani, C.; Azar, M.; Deckert, M.; Raynier, J.-L.; Clowez, G.; Boileau, P.; Ruetsch-Chelli, C. Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis. *J. Med. Internet Res.* **2024**, *26*, e53164. [CrossRef]
268. Park, S.; Nan, X. Generative AI and Misinformation: A Scoping Review of the Role of Generative AI in the Generation, Detection, Mitigation, and Impact of Misinformation. *AI Soc.* **2025**, 1–15. [CrossRef]
269. Simon, F.M.; Altay, S.; Mercier, H. Misinformation Reloaded? Fears about the Impact of Generative AI on Misinformation Are Overblown. *Harv. Kennedy Sch. Misinform. Rev.* **2023**. [CrossRef]
270. Ferrara, C.; Sellitto, G.; Ferrucci, F.; Palomba, F.; De Lucia, A. Fairness-Aware Machine Learning Engineering: How Far Are We? *Empir. Softw. Eng.* **2024**, *29*, 9. [CrossRef] [PubMed]
271. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. [CrossRef]
272. Lahusen, C.; Maggetti, M.; Slavkovik, M. Trust, Trustworthiness and AI Governance. *Sci. Rep.* **2024**, *14*, 20752. [CrossRef] [PubMed]
273. Jarmakovica, A. Machine Learning-Based Strategies for Improving Healthcare Data Quality: An Evaluation of Accuracy, Completeness, and Reusability. *Front. Artif. Intell.* **2025**, *8*, 1621514. [CrossRef]
274. Seabra, A.; Cavalcante, C.; Ruberg, N.; Lifschitz, S. *AI-Driven Semantic Data Quality Assessment and Scoring for Relational Databases*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 199–206. [CrossRef]
275. Lesouple, J.; Baudoin, C.; Spigai, M.; Tournet, J.-Y. Generalized Isolation Forest for Anomaly Detection. *Pattern Recognit. Lett.* **2021**, *149*, 109–119. [CrossRef]
276. Mohammed, S.; Budach, L.; Feuerpfeil, M.; Ihde, N.; Nathansen, A.; Noack, N.; Patzlaff, H.; Naumann, F.; Harmouch, H. The Effects of Data Quality on Machine Learning Performance on Tabular Data. *Inf. Syst.* **2025**, *132*, 102549. [CrossRef]
277. Mowla, N.I. *A Guide to Data Quality Testing for AI Applications Based on Standards*; RISE Research Institutes of Sweden: Gothenburg, Sweden, 2024.
278. EU FRA. *Data Quality and Artificial Intelligence—Mitigating Bias and Error to Protect Fundamental Rights*; European Union Agency for Fundamental Rights: Viena, Austria, 2019.
279. Pulicharla, M.R. Detecting and Addressing Model Drift: Automated Monitoring and Real-Time Retraining in ML Pipelines. *World J. Adv. Res. Rev.* **2019**, *3*, 147–152. [CrossRef]
280. Patchipala, S.G. Tackling Data and Model Drift in AI: Strategies for Maintaining Accuracy during ML Model Inference. *Int. J. Sci. Res. Arch.* **2023**, *10*, 1198–1209. [CrossRef]
281. Poppy, D. Data Governance Frameworks for AI-Driven orgs | dbt Labs. Available online: https://www.getdbt.com/blog/data-governance-frameworks-ai?utm_source=chatgpt.com (accessed on 10 November 2025).
282. Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R.B. Missing Value Estimation Methods for DNA Microarrays. *Bioinformatics* **2001**, *17*, 520–525. [CrossRef] [PubMed]
283. Li, W.; Wu, Y.; Huang, W.; Zhou, F.; Ou, W.; Wang, H.; Deng, L. System Log Anomaly Detection Based on Contrastive Learning and Retrieval Augmented. *Sci. Rep.* **2025**, *15*, 38370. [CrossRef]
284. Hansen, H.T. Intelligent Cloud-Native DevOps Architecture for Enterprise Transformation Leveraging Blockchain, BERT Models, and AI-Powered Financial Cryptosystems. *Int. J. Res. Publ. Eng. Technol. Manag.* **2025**, *8*, 1–5. Available online: <https://ijrpetm.com/index.php/IJRPETM/article/view/172/168> (accessed on 2 November 2025).
285. Meng, T.; Jing, X.; Yan, Z.; Pedrycz, W. A Survey on Machine Learning for Data Fusion. *Inf. Fusion* **2020**, *57*, 115–129. [CrossRef]
286. Ziv, L.; Nakash, M. Behind the Algorithm: International Insights into Data-Driven AI Model Development. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 122. [CrossRef]
287. Dibouliya, A. Unified Data Governance Framework for AI-Enabled Data Warehouses in Banking. *Eur. Mod. Stud. J.* **2025**, *9*, 67–76. [CrossRef]
288. Wendt, D.W. Continuous Improvement. In *AI Strategy and Security*; Apress: Berkeley, CA, USA, 2025; pp. 175–184. [CrossRef]
289. Grant, B.; Welch, M.; Deutschman, C.; McElcheran, C.; Badzynski, A.; Bell, J.A.H.; Hope, A.; Grant, R.C.; Truong, T.; Lane, K.; et al. Abstract PR-04: A Practical Framework for Operationalizing Responsible and Equitable AI in Healthcare: Tackling Bias, Inequity, and Implementation Challenges. *Clin. Cancer Res.* **2025**, *31* (Suppl. 13), PR-04. [CrossRef]
290. Bhosale, A.M. Implementing PowerBI Reporting for Quality Analysis in Decision Making Processes. Master's Thesis, Politecnico di Torino, Turin, Italy, 2025.
291. Verma, R.K. Digital Twin Technology for Process Optimization and Smart Manufacturing Systems. *Int. J. Res. Publ. Eng. Technol. Manag. (IJRPETM)* **2025**, *8*, 12699–12701.
292. Raji, I.D.; Smart, A.; White, R.N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; Barnes, P. Closing the AI Accountability Gap. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; ACM: New York, NY, USA, 2020; pp. 33–44. [CrossRef]

293. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach. (Dordr.)* **2018**, *28*, 689–707. [\[CrossRef\]](#)
294. Selbst, A.D.; Boyd, D.; Friedler, S.A.; Venkatasubramanian, S.; Vertesi, J. Fairness and Abstraction in Sociotechnical Systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; ACM: New York, NY, USA, 2019; pp. 59–68. [\[CrossRef\]](#)
295. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* **2022**, *54*, 1–35. [\[CrossRef\]](#)
296. Morley, J.; Floridi, L.; Kinsey, L.; Elhalal, A. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. In *Ethics, Governance, and Policies in Artificial Intelligence*; Floridi, L., Ed.; Springer Nature: Cham, Switzerland, 2021; pp. 144–153.
297. Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T. Model Cards for Model Reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; ACM: New York, NY, USA, 2019; pp. 220–229. [\[CrossRef\]](#)
298. Jobin, A.; Ienca, M.; Vayena, E. The Global Landscape of AI Ethics Guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [\[CrossRef\]](#)
299. Whittlestone, J.; Nyrupe, R.; Alexandrova, A.; Cave, S. The Role and Limits of Principles in AI Ethics. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*; ACM: New York, NY, USA, 2019; pp. 195–200. [\[CrossRef\]](#)
300. Goellner, S.; Tropmann-Frick, M.; Brumen, B. Responsible Artificial Intelligence: A Structured Literature Review. *arXiv* **2024**. [\[CrossRef\]](#)
301. Zeng, Y.; Lu, E.; Huangfu, C. Linking Artificial Intelligence Principles. *arXiv* **2018**. [\[CrossRef\]](#)
302. Lu, J.; Liu, A.; Dong, F.; Gu, F.; Gama, J.; Zhang, G. Learning under Concept Drift: A Review. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 2346–2363. [\[CrossRef\]](#)
303. Adepoju, A.S. Adaptive Program Management Strategies for AI-Based Cyber Defense Deployments in Critical Infrastructure and Enterprise Digital Transformation Initiatives. *Int. J. Res. Publ. Rev.* **2025**, *6*, 5599–5615. [\[CrossRef\]](#)
304. Widmer, G.; Kubat, M. Learning in the Presence of Concept Drift and Hidden Contexts. *Mach. Learn.* **1996**, *23*, 69–101. [\[CrossRef\]](#)
305. Sayles, J. Designing a Well-Governed AI Lifecycle Model. In *Principles of AI Governance and Model Risk Management*; Apress: Berkeley, CA, USA, 2024; pp. 85–111. [\[CrossRef\]](#)
306. Park, C. Addressing Challenges for the Effective Adoption of Artificial Intelligence in the Energy Sector. *Sustainability* **2025**, *17*, 5764. [\[CrossRef\]](#)
307. Mulyaningsih, S.R.; Ghoffar, A.; Mufrihah, A.; Hasanah, I.; Aun, A. Self-Adaptive Systems: Redefining Best Practices in AI and Big Data in Recruitment. In *Emerging Technologies for Recruitment Strategy and Practice*; Vasudevan, K., Vasudevan, S.K., Sudha, M., Eds.; IGI Global: Hershey, PA, USA, 2023; pp. 77–103.
308. Baum, S.D.; Owe, A. Artificial Intelligence Needs Environmental Ethics. *Ethics Policy Environ.* **2023**, *26*, 139–143. [\[CrossRef\]](#)
309. Morley, J.; Kinsey, L.; Elhalal, A.; Garcia, F.; Ziosi, M.; Floridi, L. Operationalising AI Ethics: Barriers, Enablers and next Steps. *AI Soc.* **2023**, *38*, 411–423. [\[CrossRef\]](#)
310. Asokan, D.R.; Smith, C.M.; Huq, F.A. Digitalisation as a Catalyst for Supplier Diversity, Equity and Inclusion. *Int. J. Oper. Prod. Manag.* **2025**, *2025*, 1–26. [\[CrossRef\]](#)
311. Mahler, S. Building Trust in Workplace AI Why Governance Outweighs Employee Co-Creation in Building Trust. Ph.D. Thesis, Vorarlberg University of Applied Sciences, Dornbirn, Austria, 2025.
312. Sagona, M.; Dai, T.; Macis, M.; Darden, M. Trust in AI-Assisted Health Systems and AI's Trust in Humans. *npj Health Syst.* **2025**, *2*, 10. [\[CrossRef\]](#)
313. Agate, J. Artificial Intelligence Methods and Approaches to Improve Data Quality in Healthcare Data. *Artif. Intell. Life Sci.* **2025**, *8*, 100135. [\[CrossRef\]](#)
314. Stoudt, S.; Jernite, Y.; Marshall, B.; Marwick, B.; Sharan, M.; Whitaker, K.; Danchev, V. Ten Simple Rules for Building and Maintaining a Responsible Data Science Workflow. *PLoS Comput. Biol.* **2024**, *20*, e1012232. [\[CrossRef\]](#) [\[PubMed\]](#)
315. Korbmacher, M.; Azevedo, F.; Pennington, C.R.; Hartmann, H.; Pownall, M.; Schmidt, K.; Elsherif, M.; Breznau, N.; Robertson, O.; Kalandadze, T.; et al. The Replication Crisis Has Led to Positive Structural, Procedural, and Community Changes. *Commun. Psychol.* **2023**, *1*, 3. [\[CrossRef\]](#) [\[PubMed\]](#)
316. Dudda, L.; Kormann, E.; Kozula, M.; DeVito, N.J.; Klebel, T.; Dewi, A.P.M.; Spijker, R.; Stegeman, I.; Van den Eynden, V.; Ross-Hellauer, T.; et al. Open Science Interventions to Improve Reproducibility and Replicability of Research: A Scoping Review. *R. Soc. Open Sci.* **2025**, *12*, 242057. [\[CrossRef\]](#) [\[PubMed\]](#)
317. MacMaster, S.; Sinistore, J. Testing the Use of a Large Language Model (LLM) for Performing Data Quality Assessment. *Int. J. Life Cycle Assess.* **2024**, 1–12. [\[CrossRef\]](#)

318. WHO. *Overview of the Data Quality Review (DQR) Framework and Methodology*; WHO: Geneva, Switzerland, 2020.
319. Patra, P.; Di Pompeo, D.; Di Marco, A. An Evaluation Framework for the FAIR Assessment Tools in Open Science. *arXiv* **2025**, arXiv:2503.15929. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.