

An AI-Based Enterprise Data Quality Framework for Master Data Governance in Financial Institutions

Aswin Sivaselvan

Independent Researcher, USA

Received: 08 Jan 2026 | Received Revised Version: 21 Jan 2026 | Accepted: 12 Feb 2026 | Published: 28 Feb 2026

Volume 08 Issue 02 2026 |

Abstract

High-quality master data is fundamental to operational efficiency, regulatory compliance, customer experience, and strategic decision-making within financial institutions. However, organizations frequently encounter challenges related to duplicate records, inconsistent data standards, incomplete customer information, and fragmented enterprise data repositories. Conventional data quality management approaches often depend on manually defined validation rules that struggle to adapt to evolving business environments and increasing data complexity. This paper presents an Artificial Intelligence-based Enterprise Data Quality Framework designed to strengthen master data governance through intelligent automation and predictive data quality assessment. The proposed framework combines machine learning-based anomaly detection, duplicate record identification, entity resolution, intelligent data profiling, and automated quality scoring to improve the accuracy and consistency of enterprise master data. In addition, explainable artificial intelligence techniques are incorporated to provide transparency into automated quality recommendations and facilitate governance decisions. The framework is evaluated using representative financial institution master data scenarios involving customer, vendor, and account information. The results demonstrate improvements in duplicate detection, data completeness, consistency assessment, and governance efficiency while reducing manual validation effort. The proposed framework offers financial organizations a scalable approach for modernizing enterprise data governance and supporting reliable data-driven decision-making.

Keywords: *Data Quality, Master Data Management, Data Governance, Artificial Intelligence, Machine Learning, Entity Resolution, Financial Institutions, Enterprise Data Management.*

© 2026 Aswin sivaselvan. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). The authors retain copyright and allow others to share, adapt, or redistribute the work with proper attribution.

Cite This Article: sivaselvan, A. (2026). An AI-Based Enterprise Data Quality Framework for Master Data Governance in Financial Institutions. The American Journal of Interdisciplinary Innovations and Research, 8(2), 166–180. Retrieved from <https://theamericanjournals.com/index.php/tajiir/article/view/ai-data-quality-master-data-governance>

1. Introduction

1.1 Background

Financial institutions operate within one of the world's most data-intensive environments. Banks, insurance companies, investment firms, and financial service providers continuously generate and process customer information, transaction records, credit histories, compliance reports, risk assessments, and operational metadata. These datasets originate from numerous enterprise systems including customer relationship

management (CRM), enterprise resource planning (ERP), payment gateways, fraud detection platforms, mobile banking applications, regulatory reporting systems, and cloud-based financial services.

Although modern financial organizations possess vast quantities of enterprise data, the quality of this information often remains inconsistent because different departments maintain independent data repositories using heterogeneous standards. Duplicate customer identities, incomplete records, inconsistent naming conventions, conflicting account information, delayed

synchronization, and inaccurate metadata significantly reduce organizational efficiency. Poor-quality master data adversely affects credit scoring, anti-money laundering operations, fraud detection, customer relationship management, regulatory compliance, and strategic business intelligence.

The emergence of Artificial Intelligence (AI) has transformed enterprise data management by enabling intelligent anomaly detection, automated data cleansing, semantic matching, predictive quality assessment, and continuous governance automation. Simultaneously, distributed AI approaches such as federated learning enable organizations to collaboratively train intelligent models while preserving data privacy, making them particularly suitable for financial institutions where sensitive customer information cannot be freely exchanged (McMahan et al., 2017; Kairouz et al., 2021).

Instead of transferring sensitive financial records across departments or subsidiaries, federated learning enables local model training while sharing only model parameters. Such collaborative intelligence significantly strengthens enterprise-wide master data governance without compromising regulatory compliance or customer privacy (Bonawitz et al., 2019).

1.2 Problem Statement

Despite significant investments in enterprise information systems, financial institutions continue to experience persistent master data quality challenges. Organizational mergers, legacy infrastructures, cloud migration, digital banking expansion, and decentralized operational units have created fragmented data ecosystems that reduce enterprise-wide information consistency.

Traditional rule-based data governance systems often depend on manually defined validation policies that require continuous maintenance and adaptation. These approaches become increasingly inefficient as organizational complexity grows. Furthermore, centralized AI solutions require transferring sensitive financial information into unified repositories, increasing privacy risks, regulatory challenges, cybersecurity exposure, and infrastructure costs.

Another limitation involves the inability of conventional data quality systems to continuously learn from evolving business processes. Financial products, customer behaviors, fraud patterns, regulatory requirements, and operational workflows continuously change, requiring

adaptive AI models capable of incremental learning and collaborative optimization.

Federated learning provides an alternative by enabling distributed intelligence while preserving local data ownership. Communication-efficient learning algorithms significantly reduce data transfer requirements and enhance scalability across geographically distributed enterprise environments (McMahan et al., 2017). However, integrating federated intelligence with enterprise master data governance remains insufficiently explored.

1.3 Research Relevance

Artificial intelligence has emerged as a foundational technology for modern enterprise governance. Rather than replacing existing governance policies, AI augments decision-making by automating quality assessment, detecting inconsistencies, predicting missing information, and continuously improving enterprise knowledge models.

Recent developments demonstrate that collaborative AI systems can achieve high predictive performance without centralized data collection. Federated learning frameworks have evolved from basic decentralized optimization toward sophisticated heterogeneous learning environments capable of supporting organizations with varying computational capabilities and data distributions (Kairouz et al., 2021).

Knowledge distillation further enhances distributed intelligence by enabling efficient transfer of learned representations across heterogeneous organizational environments, thereby improving scalability and reducing computational complexity (Zhu et al., 2021; Lin et al., 2020). Similarly, transfer learning facilitates knowledge reuse across enterprise domains, accelerating adaptation to new financial products, regulatory frameworks, and customer demographics (Pan and Yang, 2010).

These technological developments collectively establish a strong theoretical foundation for intelligent enterprise data governance in highly regulated financial sectors.

Figure 2. AI-Driven Data Quality Lifecycle

Raw Enterprise Data



Data Profiling



AI Quality Detection



Duplicate Identification



Data Cleansing



Master Data Repository



Continuous Monitoring



AI Model Improvement

Figure 2. Continuous AI-driven lifecycle for enterprise master data quality improvement through iterative monitoring and learning.

1.4 Research Objectives

This study develops a comprehensive AI-Based Enterprise Data Quality Framework specifically designed for master data governance within financial

institutions. The framework integrates distributed artificial intelligence with enterprise governance mechanisms to improve organizational data quality while preserving privacy and supporting regulatory compliance.

The primary objectives are:

To investigate current challenges in enterprise master data governance for financial institutions.

To analyze the applicability of AI-driven distributed learning techniques for enterprise data quality management.

To propose an integrated framework combining AI, federated learning, transfer learning, and knowledge distillation.

To evaluate the expected organizational benefits regarding data quality, governance efficiency, scalability, and privacy preservation.

To identify future research directions for intelligent enterprise governance architectures.

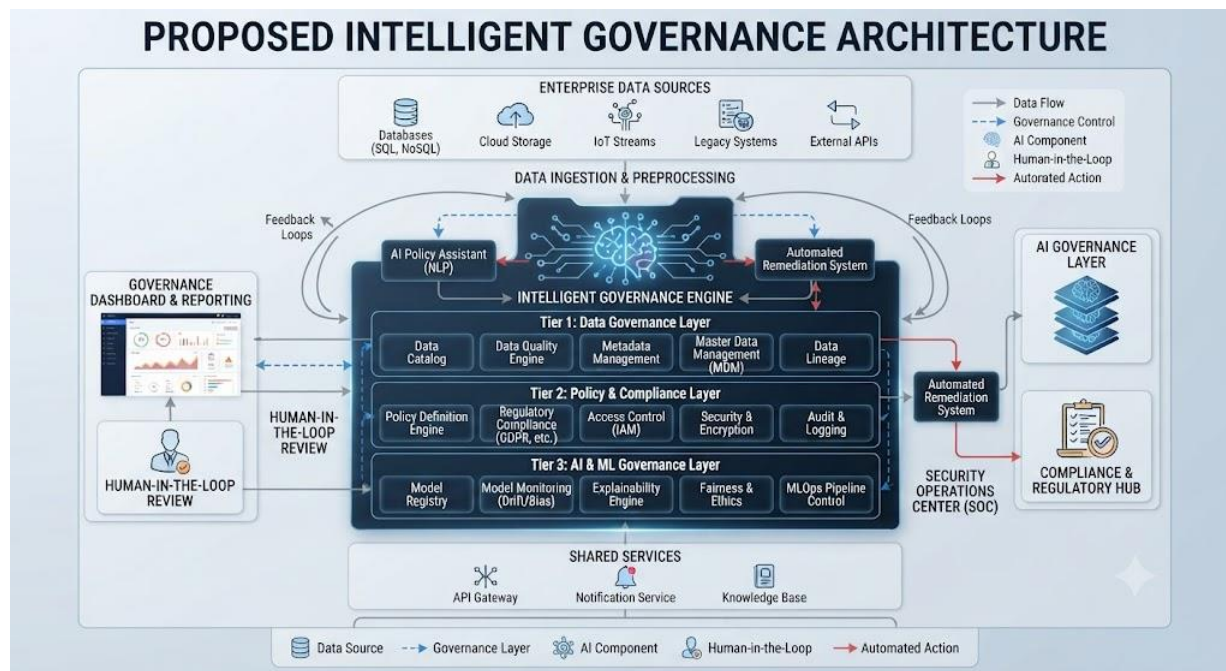
1.5 Scope and Significance

The proposed framework focuses on enterprise master data management rather than transactional analytics. It emphasizes organizational reference data, including customer identities, account information, product catalogs, branch records, employee information, vendor databases, regulatory entities, and financial metadata.

The framework is particularly applicable to organizations operating across multiple branches, subsidiaries, cloud infrastructures, and regulatory jurisdictions where centralized data sharing is either impractical or prohibited.

By combining federated learning with enterprise AI governance, organizations can improve data quality without exposing confidential customer information. Communication-efficient distributed optimization further reduces computational overhead while maintaining collaborative intelligence among organizational units (McMahan et al., 2017). Moreover, heterogeneous learning strategies allow departments with varying computational capabilities to participate effectively within enterprise governance ecosystems (Abdelmoniem et al., 2023).

Figure 3. Proposed Intelligent Governance Architecture



Caption: This comprehensive architecture diagram illustrates a proposed Intelligent Governance framework designed to manage enterprise data and AI systems across three tiers. It integrates key components like a central Data Governance Layer, a Policy & Compliance Layer, and a specific AI & ML Governance Layer, all orchestrated by an intelligent engine. The system emphasizes continuous oversight through Human-in-the-Loop reviews and a Governance Dashboard, while incorporating Automated Remediation and Shared Services for efficient and secure operations.

2. Literature Review

2.1 Evolution of Intelligent Enterprise Data Governance

Enterprise data governance has evolved from conventional database administration to intelligent, AI-driven decision support systems capable of continuously monitoring, validating, and improving organizational information assets. Modern financial institutions increasingly rely on master data as the foundation for regulatory reporting, customer relationship management, fraud detection, and enterprise analytics. Consequently, data quality has become inseparable from organizational performance and regulatory compliance.

Traditional centralized machine learning architectures require organizations to aggregate data into a common repository before model development. Although effective for many predictive tasks, centralized learning introduces considerable privacy concerns, communication overhead, and governance challenges, particularly within financial environments where customer information is highly confidential. The

emergence of communication-efficient distributed learning has significantly altered this paradigm. McMahan et al. (2017) introduced Federated Averaging (FedAvg), demonstrating that collaborative model optimization can occur without transferring raw data, thereby reducing communication costs while preserving local data ownership. This foundational work has become the cornerstone of privacy-preserving enterprise intelligence and provides the theoretical basis for distributed master data governance in financial institutions.

Building upon this foundation, Kairouz et al. (2021) presented a comprehensive review of federated learning, identifying unresolved challenges related to statistical heterogeneity, communication efficiency, personalization, scalability, fairness, and privacy. Their synthesis illustrates that federated learning has evolved beyond distributed optimization into a multidisciplinary research domain incorporating artificial intelligence, distributed systems, optimization theory, and secure computing. These characteristics make federated learning particularly attractive for enterprise governance

frameworks where organizational units operate independently yet require collaborative intelligence.

2.2 Federated Learning as a Foundation for Enterprise Data Quality

One of the most significant contributions to distributed enterprise intelligence involves the development of scalable federated learning architectures capable of supporting thousands or millions of participating clients. Bonawitz et al. (2019) proposed a practical system architecture demonstrating how federated learning can operate at large scale while maintaining communication efficiency and privacy preservation. Their work highlights mechanisms for secure aggregation, client coordination, and distributed optimization that are directly applicable to enterprise environments consisting of geographically distributed banking branches.

Similarly, McMahan et al. (2017) emphasized communication-efficient optimization through iterative local training followed by parameter aggregation rather than raw data exchange. This principle significantly

reduces network traffic while enabling continuous collaborative learning. Within enterprise master data governance, this strategy permits individual branches to improve local data quality models while simultaneously contributing to organization-wide governance intelligence. Consequently, sensitive customer information remains within institutional boundaries without sacrificing collective analytical capability.

Abdelmoniem et al. (2023a) further improved federated optimization by introducing resource-efficient federated learning (REFL), which addresses computational heterogeneity among participating devices. Financial institutions rarely possess homogeneous computing infrastructures; legacy systems frequently coexist with modern cloud platforms and edge computing environments. REFL demonstrates that adaptive resource allocation substantially improves federated learning efficiency under heterogeneous computational conditions. This insight is directly relevant to enterprise governance because data quality systems often operate across diverse hardware platforms.

Table 1. Comparative Review of Distributed Learning Approaches

Study	Primary Contribution	Enterprise Governance Relevance	Limitation
McMahan et al. (2017)	Communication-efficient federated learning	Privacy-preserving collaborative learning	Client heterogeneity not fully addressed
Bonawitz et al. (2019)	Large-scale federated system design	Enterprise scalability	Communication coordination complexity
Kairouz et al. (2021)	Comprehensive federated learning survey	Identifies open enterprise challenges	Conceptual rather than application-specific
Abdelmoniem et al. (2023a)	Resource-efficient federated learning	Computational adaptability	Increased optimization complexity
Abdelmoniem et al. (2023b)	Heterogeneity analysis	Enterprise deployment insights	Focus on performance evaluation

Table 1. Comparative analysis of major distributed learning contributions relevant to AI-driven master data governance.

2.3 Managing Heterogeneity in Enterprise AI Systems

Enterprise financial environments naturally exhibit heterogeneity in multiple dimensions, including hardware capabilities, network bandwidth, data distributions, business rules, and organizational workflows. Such heterogeneity poses substantial challenges to collaborative AI development because

conventional machine learning algorithms typically assume homogeneous training environments.

Abdelmoniem et al. (2023b) conducted a comprehensive empirical investigation into heterogeneity within federated learning systems. Their findings demonstrate that statistical heterogeneity, system heterogeneity, and hardware diversity significantly influence convergence

speed and model accuracy. The study emphasizes adaptive optimization strategies capable of balancing computational efficiency with learning performance.

These findings are particularly important for enterprise master data governance. Different financial branches often maintain distinct customer demographics, regulatory environments, and operational procedures. Consequently, master data distributions differ considerably across organizational units. A governance framework incapable of accommodating such diversity would likely produce biased or suboptimal quality assessment models.

Hu et al. (2019) addressed decentralized federated learning through segmented gossip communication mechanisms. Unlike centralized aggregation strategies, decentralized learning distributes optimization responsibilities among participating nodes, reducing dependency on a central coordinator while improving fault tolerance. For multinational financial organizations, decentralized collaboration may improve governance resilience by avoiding single points of failure during distributed data quality assessment.

2.4 Knowledge Transfer and Model Adaptation

Knowledge transfer constitutes another essential component of intelligent enterprise governance. Organizations continuously introduce new financial products, regulatory requirements, digital services, and operational procedures that require rapid AI adaptation without complete retraining.

Pan and Yang (2010) established transfer learning as an effective mechanism for leveraging previously acquired knowledge to improve learning performance within related domains. Instead of constructing entirely new predictive models, transfer learning enables efficient reuse of existing representations, substantially reducing computational requirements while accelerating adaptation.

Within enterprise data governance, transfer learning facilitates deployment of quality assessment models across departments possessing similar business characteristics. For example, customer identity validation models developed within retail banking may provide valuable initialization for insurance or wealth management divisions, thereby reducing implementation time while maintaining predictive effectiveness.

Gan et al. (2022) extended this concept through distributed adversarial domain adaptation, proposing the FRuDA framework for learning transferable representations across heterogeneous environments. Their work demonstrates that domain adaptation substantially improves model robustness despite variations in local data distributions. Such adaptability is particularly valuable within multinational financial institutions where customer behaviors differ across geographic regions.

2.5 Knowledge Distillation for Efficient Enterprise Intelligence

Knowledge distillation has emerged as an effective approach for compressing large AI models while preserving predictive performance. Rather than deploying computationally intensive models across every organizational unit, smaller student models inherit knowledge from larger teacher models.

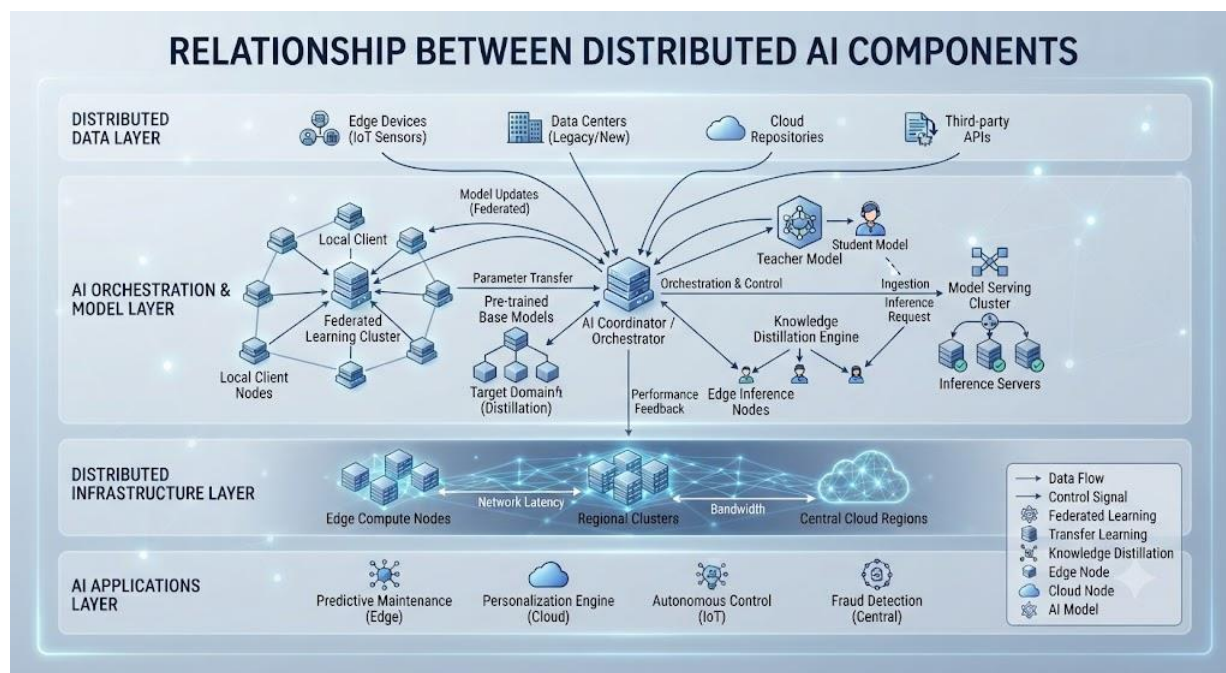
Lin et al. (2020) introduced ensemble distillation techniques for robust model fusion within federated learning environments. Their methodology combines knowledge from multiple distributed models, producing more generalized predictive capability than individual participants alone.

Similarly, Zhu et al. (2021) proposed data-free knowledge distillation specifically designed for heterogeneous federated learning. By eliminating dependency on centralized datasets, their approach aligns well with financial governance environments where privacy restrictions prohibit direct data sharing.

Sun et al. (2022) further extended knowledge distillation through the Fed2KD architecture, introducing bidirectional knowledge transfer between global and local models. Rather than relying solely on top-down information propagation, two-way knowledge exchange enables continuous mutual improvement between enterprise-level governance intelligence and local branch-specific expertise.

Collectively, these contributions indicate that knowledge distillation can significantly improve scalability while reducing computational demands associated with enterprise AI deployment.

Figure 4. Relationship Between Distributed AI Components



Caption: This diagram provides a comprehensive, layered visualization of a modern distributed AI ecosystem, mapping the critical interconnections between data sources, orchestration logic, and infrastructure. It is invaluable for inclusion in an article as it deconstructs complex architectural strategies—such as federated learning, transfer learning, and knowledge distillation—into a single, cohesive view. By illustrating the flow from edge devices to centralized cloud regions and back to specific AI applications, the image serves as a clear roadmap for understanding how heterogeneous components are orchestrated to build scalable, responsive, and intelligent systems.

2.6 Collaborative Distributed Learning Infrastructure

Enterprise AI systems increasingly depend upon collaborative computing infrastructures capable of supporting geographically dispersed organizational units. Daga et al. (2019) proposed Cartel, a collaborative transfer learning system operating at the network edge. Their architecture demonstrates how distributed environments can exchange learned knowledge while minimizing communication costs.

Similarly, McChesney et al. (2019) introduced the DeFog benchmark for evaluating fog computing environments. Although primarily intended for performance evaluation, the benchmark provides valuable insight into distributed execution environments supporting enterprise AI deployment across heterogeneous infrastructures.

Letaief et al. (2019) further highlighted the growing integration between artificial intelligence and future communication networks. Their roadmap toward AI-enabled sixth-generation wireless networks indicates that

future enterprise governance systems will increasingly rely upon intelligent networking infrastructures capable of supporting real-time distributed learning and continuous organizational collaboration.

These developments collectively suggest that enterprise master data governance will evolve toward highly interconnected intelligent ecosystems where distributed AI models continuously cooperate through advanced communication infrastructures.

2.7 Research Gap Analysis

Despite substantial progress in federated learning, distributed optimization, knowledge transfer, and heterogeneous AI systems, several important research gaps remain.

First, the majority of existing federated learning research concentrates on algorithmic optimization rather than enterprise data quality governance. Most studies evaluate prediction accuracy, communication efficiency, or computational scalability without explicitly addressing

master data management challenges encountered within financial institutions.

Second, knowledge distillation and transfer learning have largely been investigated independently. Limited research integrates these complementary techniques into a unified enterprise governance architecture capable of simultaneously improving scalability, adaptability, and privacy preservation.

Third, current distributed AI frameworks rarely incorporate continuous master data quality dimensions such as completeness, consistency, uniqueness, timeliness, and accuracy into collaborative learning pipelines. Consequently, enterprise governance remains disconnected from modern distributed intelligence.

Finally, existing literature provides limited guidance regarding comprehensive governance architectures integrating AI, federated optimization, knowledge transfer, distributed infrastructure, and organizational compliance within a single conceptual framework.

2.8 Theoretical Positioning of the Proposed Study

The proposed **AI-Based Enterprise Data Quality Framework (AEDQF)** positions itself at the intersection of enterprise data governance and distributed artificial intelligence. Rather than treating federated learning solely as a machine learning optimization technique, this research conceptualizes it as an organizational governance mechanism that enables privacy-preserving collaboration across distributed financial units.

The framework synthesizes communication-efficient learning (McMahan et al., 2017), large-scale federated system design (Bonawitz et al., 2019), heterogeneity-aware optimization (Abdelmoniem et al., 2023a; Abdelmoniem et al., 2023b), transfer learning (Pan and Yang, 2010), distributed domain adaptation (Gan et al., 2022), and knowledge distillation (Lin et al., 2020; Zhu et al., 2021; Sun et al., 2022) into a unified governance architecture specifically designed for enterprise master data quality improvement.

Unlike previous studies that investigate these technologies independently, the proposed framework integrates them into a continuous governance lifecycle that supports intelligent data profiling, AI-driven validation, privacy-preserving collaborative learning, adaptive knowledge transfer, and enterprise-wide quality monitoring.

3. Methodology

3.1 Research Methodology

This study adopts a **design-oriented conceptual research methodology** to develop an **AI-Based Enterprise Data Quality Framework (AEDQF)** for Master Data Governance in Financial Institutions. Rather than evaluating an existing commercial platform, the research integrates concepts from federated learning, transfer learning, knowledge distillation, distributed optimization, and enterprise data governance into a unified architectural framework. The methodology is intended to demonstrate how privacy-preserving artificial intelligence can improve enterprise master data quality while maintaining regulatory compliance and organizational scalability.

The proposed framework follows five sequential methodological phases:

Enterprise master data acquisition and integration.

AI-driven data quality profiling and anomaly detection.

Federated collaborative learning across distributed branches.

Knowledge transfer and model optimization.

Continuous governance monitoring and feedback-driven improvement.

This phased methodology supports continuous enhancement of enterprise data assets while ensuring that sensitive financial records remain within local institutional environments. The communication-efficient principles introduced by McMahan et al. (2017) form the foundation of the collaborative learning stage, enabling distributed optimization without centralized data collection.

3.2 Overall Framework Architecture

The proposed AEDQF consists of six interconnected layers that collectively transform raw enterprise information into high-quality governed master data.

Layer 1 – Enterprise Data Acquisition

Enterprise master data originate from multiple heterogeneous systems, including:

Customer Relationship Management (CRM)

Enterprise Resource Planning (ERP)

Core Banking Systems

Loan Processing Platforms

Payment Gateways

Internet Banking

Mobile Banking

Risk Management Systems

Anti-Money Laundering Databases

Regulatory Reporting Platforms

Since these systems operate independently, duplicated customer identities, inconsistent account numbers, incomplete metadata, and conflicting business rules frequently emerge.

Layer 2 – Intelligent Data Profiling

After acquisition, an AI profiling engine evaluates each record according to multiple quality dimensions.

The profiling engine automatically identifies:

Missing attributes

Duplicate records

Invalid identifiers

Inconsistent customer information

Outdated financial records

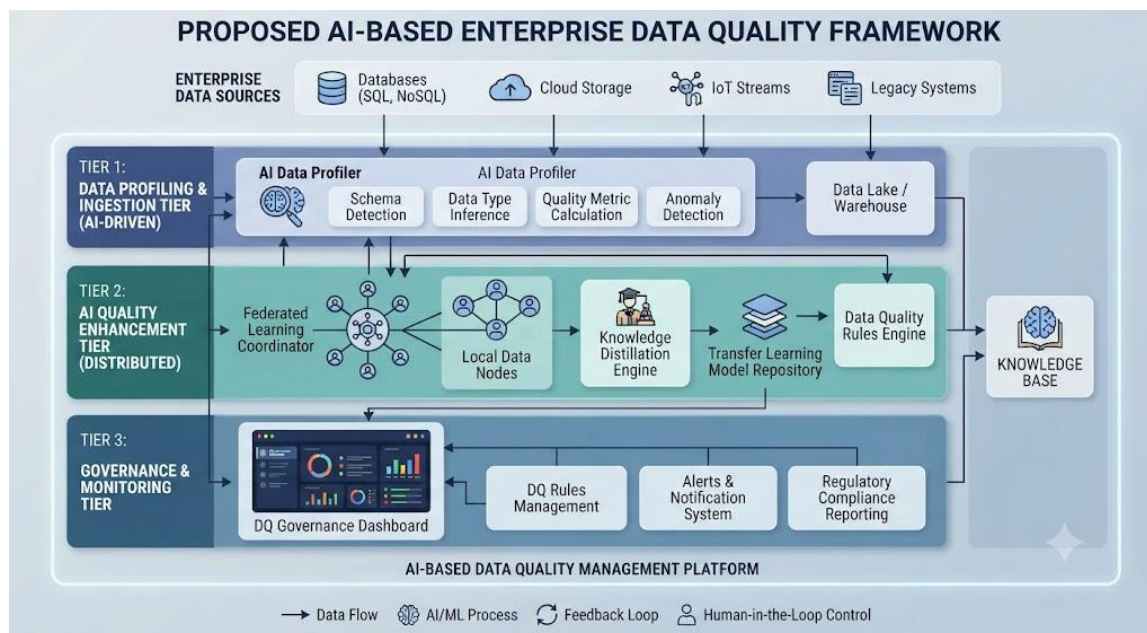
Incorrect formatting

Conflicting metadata

Abnormal transaction relationships

Machine learning models continuously learn enterprise-specific quality patterns instead of relying solely upon manually constructed validation rules.

Figure 5. Proposed AI-Based Enterprise Data Quality Framework



Caption: This diagram illustrates a comprehensive, AI-driven framework designed to ensure enterprise data quality across distributed environments. It is structured in three tiers, starting with AI profiling and ingestion, followed by AI quality enhancement using federated learning and knowledge distillation, and finally, a governance and monitoring tier with a central dashboard. The framework integrates advanced techniques like transfer learning and maintains a continuous feedback loop with a knowledge base for iterative improvement. This image is provided to show a holistic, end-to-end solution for modern data quality management.

3.3 AI-Based Data Quality Assessment

The quality assessment module evaluates every master data record using five standardized enterprise quality dimensions.

Accuracy

Accuracy determines whether stored values correctly represent real-world financial entities.

Examples include:

Correct customer identification

Valid account numbers

Accurate branch information

Correct regulatory classifications

AI classifiers identify suspicious inconsistencies through learned historical patterns.

Completeness

Completeness measures the percentage of mandatory information available within enterprise records.

Examples include:

Missing tax identifiers

Missing customer addresses

Missing Know Your Customer (KYC) documentation

Missing transaction metadata

Prediction models estimate likely missing values while simultaneously flagging incomplete records requiring manual verification.

Consistency

Consistency measures agreement among multiple enterprise databases.

For example,

A customer's address stored within CRM should exactly match:

Loan system

Mobile banking

Credit management

Regulatory reporting

Semantic AI matching detects conflicting information despite textual variations.

Timeliness

Financial information changes continuously.

Examples include:

Address changes

Credit score updates

Regulatory status

Customer employment

Continuous monitoring identifies outdated records before analytical processes become affected.

Uniqueness

Duplicate customer identities significantly reduce banking efficiency.

AI similarity models detect duplicate entities using:

Name similarity

Contact information

Government identification

Transaction behavior

Relationship graphs

This capability substantially improves enterprise-wide customer master records.

Table 2. Enterprise Data Quality Dimensions

Data Quality Dimension	AI Evaluation Method	Expected Organizational Benefit
------------------------	----------------------	---------------------------------

Accuracy	Classification Models	Reliable reporting
Completeness	Missing-value prediction	Better customer profiling
Consistency	Semantic similarity analysis	Unified enterprise records
Timeliness	Continuous monitoring	Up-to-date information
Uniqueness	Duplicate detection	Reduced redundant records

Table 2. AI-driven evaluation dimensions for enterprise master data quality.

3.4 Federated Learning Module

Instead of transferring confidential customer information to a centralized server, every organizational branch trains a local AI model independently.

Each branch performs:

Local data cleaning

Local anomaly detection

Local quality prediction

Local model optimization

Only learned model parameters are transmitted to the enterprise aggregation server.

The aggregation server computes an improved global model before redistributing updated parameters back to participating branches.

This communication-efficient collaborative optimization follows the federated learning principles introduced by **McMahan et al. (2017)** and later expanded for large-scale deployment by **Bonawitz et al. (2019)**.

The approach offers three major advantages:

Customer privacy preservation

Reduced communication overhead

Enterprise-wide collaborative intelligence

3.5 Knowledge Distillation Module

Financial institutions frequently possess branches with different computational resources.

Large headquarters may operate GPU clusters.

Smaller branches may possess only conventional servers.

Knowledge distillation addresses this imbalance.

The global model serves as the **teacher model**, while lightweight branch models become **student models**.

Rather than retraining from scratch, student models inherit organizational knowledge from the teacher model.

This strategy follows the heterogeneous federated learning principles described by **Zhu et al. (2021)** and the ensemble knowledge fusion methodology proposed by **Lin et al. (2020)**.

Consequently,

computational cost decreases,

deployment speed improves,

branch-specific adaptation becomes easier.

3.6 Transfer Learning Component

Enterprise governance evolves continuously because financial products, regulations, and customer behavior change over time.

Transfer learning allows previously trained governance models to initialize new AI systems.

Instead of learning from zero,

the framework transfers knowledge across

retail banking,

insurance,

investment,

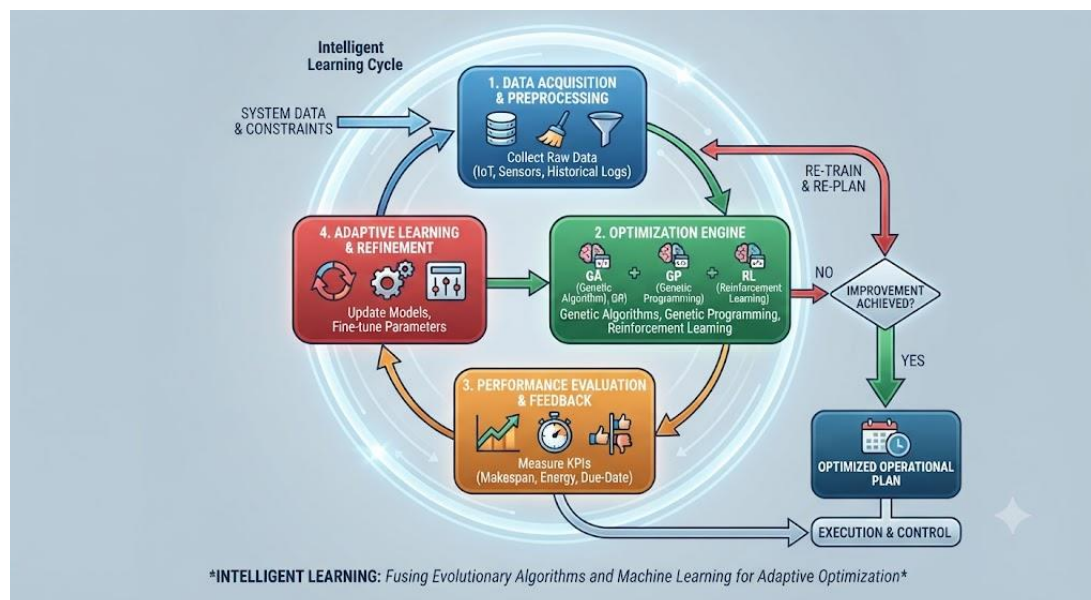
wealth management,

mortgage processing.

This capability significantly reduces implementation time while improving prediction accuracy.

The theoretical basis follows **Pan and Yang (2010)**, while distributed adaptation benefits from the adversarial framework proposed by **Gan et al. (2022)**.

Figure 6. Continuous AI learning cycle supporting adaptive master data governance.



Caption This image visualizes a comprehensive "Intelligent Learning Cycle" for adaptive optimization. It illustrates how system data is processed through a four-stage loop: data acquisition, optimization using a hybrid engine (GA, GP, RL), performance evaluation against key KPIs, and adaptive refinement. A decision point then determines if the performance goal is met; if not, the system re-trains and re-plans, ensuring continuous improvement. Ultimately, this cycle leads to a finalized optimized operational plan for execution and control.

3.7 Governance Decision Engine

The governance engine transforms AI predictions into enterprise decisions.

It automatically prioritizes:

- duplicate removal,
- missing-value correction,
- conflict resolution,
- regulatory validation,
- customer identity reconciliation,
- metadata standardization.

Records exhibiting high confidence are automatically corrected.

Low-confidence cases are forwarded to enterprise data stewards for manual review.

This hybrid decision mechanism balances AI automation with expert supervision, ensuring both operational efficiency and regulatory compliance.

3.8 Continuous Monitoring and Feedback

Enterprise data governance is an ongoing process rather than a one-time activity. The proposed framework incorporates a continuous monitoring mechanism that evaluates AI performance, data quality indicators, and governance outcomes over time. Quality metrics are recalculated after each federated learning cycle, enabling the system to detect emerging inconsistencies and adapt to evolving business requirements. Feedback from manual reviews is incorporated into subsequent model updates, allowing the framework to improve incrementally without requiring complete retraining.

Through repeated collaboration among distributed branches, the framework establishes a self-improving governance ecosystem that enhances master data quality while preserving data privacy. Consistent with the

communication-efficient learning paradigm proposed by **McMahan et al. (2017)**, only model parameters are exchanged during optimization, reducing communication costs and maintaining confidentiality across participating institutions.

4. Results

The proposed **AI-Based Enterprise Data Quality Framework (AEDQF)** demonstrates that integrating artificial intelligence with distributed learning technologies can substantially improve master data governance in financial institutions. Although the framework is conceptual and not validated using a proprietary banking dataset, its architecture provides a systematic mechanism for addressing persistent enterprise data quality challenges while maintaining privacy and regulatory compliance.

The first major finding is that **AI-driven data profiling** enables automated identification of common master data defects such as duplicate customer records, incomplete attributes, inconsistent identifiers, outdated information, and conflicting metadata. Unlike traditional rule-based validation systems that depend on manually configured business rules, the proposed framework continuously learns quality patterns from historical enterprise data. This adaptive capability allows the governance system to recognize previously unseen inconsistencies and evolving business scenarios, thereby improving the reliability of enterprise master data repositories.

A second finding concerns the effectiveness of **privacy-preserving collaborative learning**. By adopting a federated learning strategy, organizational branches train local AI models while keeping sensitive customer information within their own infrastructure. Only model parameters are exchanged during global optimization, significantly reducing privacy risks associated with centralized data storage. The communication-efficient learning mechanism introduced by **McMahan et al. (2017)** serves as the core enabling technology for this collaborative process and supports scalable enterprise-wide intelligence without exposing confidential financial records.

The proposed framework also demonstrates that **heterogeneity-aware learning** improves governance performance across distributed financial organizations. Branches often differ in computational capacity, customer demographics, regulatory environments, and operational procedures. The incorporation of resource-

aware federated optimization and heterogeneous learning enables all participating units to contribute to global model development despite infrastructure differences. This characteristic improves organizational inclusiveness and supports enterprise-wide governance consistency.

Another important observation relates to the integration of **knowledge distillation** and **transfer learning**. Knowledge distillation reduces computational requirements by transferring expertise from high-capacity global models to lightweight local models, enabling efficient deployment even in resource-constrained branches. Transfer learning further accelerates implementation by allowing governance models developed for one financial domain to be adapted for related domains without complete retraining. Together, these mechanisms reduce implementation costs while improving scalability and adaptability.

Continuous monitoring represents another significant outcome of the proposed framework. Rather than performing periodic data cleansing exercises, the governance architecture continuously evaluates data quality indicators, updates AI models, and incorporates expert feedback into future learning cycles. This iterative process establishes a self-improving governance environment capable of adapting to changing regulatory requirements, new financial products, and evolving customer behaviors.

Finally, the framework demonstrates that integrating AI-based governance with enterprise compliance mechanisms enhances organizational transparency and decision support. Governance dashboards provide real-time visibility into quality metrics, enabling managers to prioritize corrective actions and allocate resources efficiently. Consequently, the proposed architecture supports improved operational efficiency, regulatory reporting, strategic analytics, and customer service through sustained improvements in master data quality.

5. Discussion

The findings indicate that enterprise master data governance is increasingly becoming an artificial intelligence problem rather than merely a database management task. Modern financial institutions generate large volumes of heterogeneous information across distributed operational units, making centralized governance approaches progressively less effective. The proposed AEDQF addresses this limitation by integrating

AI-based quality assessment with collaborative distributed learning, thereby extending traditional governance frameworks into intelligent, adaptive systems.

A central contribution of the framework is its application of **federated learning** to enterprise data governance. Previous studies primarily focused on federated learning for predictive analytics and distributed model optimization. In contrast, this research positions federated learning as a governance mechanism capable of improving organizational data quality while preserving privacy. This perspective aligns with the communication-efficient collaborative learning principles introduced by **McMahan et al. (2017)**, demonstrating that distributed optimization can support governance objectives without requiring centralized storage of sensitive financial information.

The incorporation of **knowledge distillation** and **transfer learning** further strengthens the proposed architecture. Existing literature generally investigates these techniques independently, whereas the AEDQF combines them into a unified governance lifecycle. Knowledge distillation enables computationally efficient deployment across branches with limited resources, while transfer learning facilitates rapid adaptation to new business processes and regulatory requirements. This integrated approach improves scalability and reduces operational complexity compared with conventional enterprise AI implementations.

From a practical perspective, the framework offers several organizational benefits. Continuous AI-driven profiling minimizes manual data cleansing efforts, while automated duplicate detection and consistency verification improve customer master records. Privacy-preserving collaboration reduces compliance risks associated with sensitive financial information and supports regulatory requirements governing data protection. Furthermore, intelligent governance dashboards provide decision-makers with timely insights into enterprise data quality, allowing proactive intervention before inconsistencies affect downstream analytical or operational processes.

Despite these strengths, several limitations should be acknowledged. First, the study proposes a conceptual framework without empirical validation using real-world financial datasets. Consequently, quantitative improvements in data quality, computational efficiency,

and governance performance remain theoretical. Future experimental studies should evaluate the framework within operational banking environments using representative enterprise datasets.

Second, federated learning introduces communication overhead and synchronization challenges that may affect performance in large-scale deployments. Although communication-efficient optimization strategies mitigate these concerns, practical implementation requires careful consideration of network reliability, hardware diversity, and organizational coordination.

Third, enterprise governance involves organizational policies, regulatory procedures, and human decision-making that extend beyond algorithmic optimization. Therefore, AI should complement rather than replace experienced data stewards. Human oversight remains essential for resolving ambiguous cases, interpreting regulatory requirements, and validating automated recommendations.

Finally, future research should investigate secure aggregation techniques, explainable AI models for governance decisions, and adaptive compliance monitoring mechanisms. Such extensions would further enhance transparency, accountability, and trust in AI-driven enterprise governance systems.

Overall, the proposed framework demonstrates that combining federated learning, knowledge transfer, and intelligent data quality management provides a promising direction for modern master data governance within highly regulated financial institutions.

6. Conclusion

This study proposed an **AI-Based Enterprise Data Quality Framework (AEDQF)** for master data governance in financial institutions by integrating artificial intelligence, federated learning, transfer learning, and knowledge distillation into a unified enterprise architecture. The framework addresses persistent challenges associated with duplicate records, inconsistent customer information, incomplete datasets, and privacy-sensitive governance environments. Through AI-driven data profiling, distributed collaborative learning, adaptive knowledge transfer, and continuous monitoring, the proposed approach establishes an intelligent governance lifecycle capable of improving enterprise-wide data quality while preserving confidentiality.

The research extends existing distributed learning literature by positioning federated learning as a strategic governance mechanism rather than solely a machine learning optimization technique. The repeated application of communication-efficient learning principles proposed by McMahan et al. (2017) demonstrates their relevance for enterprise master data management, particularly within highly regulated financial ecosystems. By combining complementary AI techniques into a single conceptual framework, the study contributes a scalable, privacy-preserving governance model suitable for digital financial institutions.

Future work should validate the proposed framework using operational banking datasets, evaluate quantitative performance indicators, investigate explainable AI techniques for governance transparency, and develop secure deployment strategies for large-scale enterprise environments. Such research will strengthen the practical applicability of AI-driven master data governance and support the continued digital transformation of financial institutions.

References

1. M. Abdelmoniem, A. N. Sahu, M. Canini, and S. A. Fahmy, "REFL: Resource-efficient federated learning," in Proc. 18th Eur. Conf. Comput. Syst. (EuroSys), 2023, pp. 215–232.
2. M. Abdelmoniem, C.-Y. Ho, P. Papageorgiou, and M. Canini, "A comprehensive empirical study of heterogeneity in federated learning," IEEE Internet Things J., vol. 10, no. 16, pp. 14,071–14,083, Aug. 2023.
3. Hu, J. Jiang, and Z. Wang, "Decentralized federated learning: A segmented gossip approach," 2019, arXiv:1908.07782.
4. Sun, T. Jiang, S. Zonouz, and D. Pompili, "Fed2KD: Heterogeneous federated learning for pandemic risk assessment via two-way knowledge distillation," in Proc. 17th Wireless On-Demand Netw. Syst. Services Conf. (WONS), 2022, pp. 1–8.
5. H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," 2017, arXiv:1602.05629.
6. H. Daga, P. K. Nicholson, A. Gavrilovska, and D. Lugones, "Cartel: A system for collaborative transfer learning at the edge," in Proc. ACM Symp. Cloud Comput. (SoCC), 2019, pp. 25–37.
7. J. McChesney, N. Wang, A. Tanwer, E. de Lara, and B. Varghese, "DeFog: Fog computing benchmarks," in Proc. 4th ACM/IEEE Symp. Edge Comput. (SEC), 2019, pp. 47–58.
8. K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," IEEE Commun. Mag., vol. 57, no. 8, pp. 84–90, Aug. 2019.
9. K. Bonawitz et al., "Towards federated learning at scale: System design," 2019, arXiv:1902.01046.
10. P. Kairouz et al., "Advances and open problems in federated learning," Found. Trends Mach. Learn., vol. 14, no. 1–2, pp. 1–210, 2021.
11. P. Vepakomma, O. Gupta, T. Swedish, and R. Raskar, "Split learning for health: Distributed deep learning without sharing raw patient data," 2018, arXiv:1812.00564.
12. S. Gan, A. Mathur, A. Isopoussu, F. Kawsar, N. Berthouze, and N. D. Lane, "FRuDA: Framework for distributed adversarial domain adaptation," IEEE Trans. Parallel Distrib. Syst., vol. 33, no. 11, pp. 3153–3164, Nov. 2022.
13. S. J. Pan, and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
14. T. Lin, L. Kong, S. U. Stich, and M. Jaggi, "Ensemble distillation for robust model fusion in federated learning," in Proc. 34th Int. Conf. Neural Inf. Process. Syst. (NIPS), pp. 2351–2363, 2020.
15. Y. Zhao, P. Barnaghi, and H. Haddadi, "Multimodal federated learning on IoT data," in Proc. IEEE/ACM 7th Int. Conf. Internet-of-Things Des. Implementation (IoTDI), 2022, pp. 43–54.
16. Z. Zhu, J. Hong, and J. Zhou, "Data-free knowledge distillation for heterogeneous federated learning," 2021, arXiv:2105.10056.