

Research on Financial Data Governance Framework and Data Quality Evaluation in the Big Data Environment

Jiaqing Zhang

Small Business Banking, Capital one, Mclean, Virginia, 22102, US

Keywords: Big data; financial data governance; Data quality evaluation; Data governance framework; Financial digitalization

Abstract: Financial Data is becoming a vital part of the financial institution's process when it comes to making credit decisions, monitoring transactions, regulatory reporting and business analysis. With the mounting volume of customer data, account data, transaction data, risk data and external market data, financial data governance issues have become more apparent, such as inconsistent standards, fluctuating data quality, disjointed data systems, and increased compliance requirements. Literature review and inductive analysis are used in this study to analyze the financial data governance framework and financial data quality evaluation. The authors conclude the study by asserting that financial data governance should encompass data access, standard unification, storage integration, data quality control, compliance security and application feedback. The stages are designed to solve issues like data sources being spread across, indicator definitions not being consistent, missing and duplicated data, and challenges with data use across systems and unclear security responsibilities. Data quality evaluation shouldn't just be about scoring results. Rather, it should focus on a variety of business use cases like credit approval, fraud detection, risk monitoring, and regulatory reporting, and should measure data quality across the dimensions of completeness, accuracy, consistency, timeliness, uniqueness, security, and traceability. The study identifies that financial data governance is still poorly coordinated between governance processes and quality evaluation indicators, as well as there are few evaluation indicators based on scenarios and feedback mechanisms are not complete. All financial future data governance should focus more on data issues in business applications, so that the results of data quality evaluation can be the basis for continuous improvement of data standard and system interface and data responsibility mechanism and data security management.

1. Introduction

Financial institutions are becoming data-driven more and more day-by-day. It is required to have fairly accurate and complete data for customer credit assessment, transaction monitoring, risk warning and regulatory reporting. A vast amount of customer information, account data, transaction records, credit records, risk indicators and external market data are generated continuously by banks,

securities firms, insurance companies and internet finance platforms. Financial data has been facilitated by big data technologies to better process and analyse the data, which helps in assessing risk, evaluating businesses and taking decisions [1].

The more data that there is, the more data governance becomes challenging. Field names, statistical definitions, data format and update frequencies vary between different business systems. There are also close ties to financial data for credit approval, fraud detection, financial auditing, risk monitoring, and regulatory reporting. Inaccurate, incomplete, late or inconsistent data can impact business decisions and could even exacerbate financial risks. Although many studies have pointed out that financial auditing and risk identification are transforming in the big data era, the challenges of incomplete data standards, insufficient data quality control, and unclear responsibility boundaries in data collection, processing, sharing and application are still existing in financial institutions [2]. Demands for data authenticity, completeness, compliance and traceability in national auditing and financial regulations are also rising. Financial data governance is therefore a technical matter, as well as institutional, organizational and regulatory coordination [3].

The current research mostly focuses on financial data governance through the lenses of data standards, data quality, security compliance, risk prevention and data finance. Some studies claim that stable data quality is required to perform financial innovation, and the high-quality, complete, consistent, and timely data can help in financial modeling, risk monitoring, and intelligent decision making [4]. Studies on financial big data modeling also reveal that, besides the algorithms, the quality of raw data sources, the completeness of samples, the validity of the features and the bias of the data all affect the performance of the models. Researches on the layers of financial data governance also indicate that financial data governance should not be at one level. Rather, resources and data, governance policies, business, regulatory needs, and technical assistance must be aligned [6].

In this context, data collection or data cleaning only discussions do not suffice to encompass the entire financial data governance landscape. Financial institutions must look at all aspects of the process, such as data access, data standardization, data storage integration, data quality management, security compliance and application services. Further, the roles of governance for each stage should be explained and how they relate to business scenarios.

Data quality assessment is one of the key means to measure the value of data governance. Financial data can't be evaluated solely by the size of the data or amount of data stored; it's important to consider whether the data can support business requirements. For instance, customer identification data and credit approval are impacted by missing customer identity data. Transaction data may be delayed, which may impact response time for fraud detection. When a system of indicating inconsistencies, it can impact regulation reporting and business analysis. Previous studies show that financial data quality assessment should not only include the financial data itself, but also the operation of the platform, the stability of financial services and quality management mechanisms [7]. Thus, data quality assessment should begin with the business needs, and then develop an assessment system that incorporates categories like completeness, accuracy, consistency, timeliness, uniqueness, security, and traceability, and then incorporate the results of the data quality assessment into the data governance process.

While some of the prior work has touched on the financial data governance issues from an institutional, technical and risk control point of view, there is still some work to be done on the relationship between data governance and data quality assessment. Other studies are more oriented towards the macro level institutional analysis and fail to discuss the linkage between stages of internal governance in financial institutions. Still others focus more on the technical aspects of processing and quality control, neglecting organizational responsibility, the mechanisms of processes, and regulatory requirements. Often, data quality evaluation research is limited to

indicator listing and the application of one method, with a lack of discussion on the applicability of the method, limitations and the feedback of the results to the governance practices. Financial data governance is also found to have issues in accordance with related studies, including barriers to data, lack of unified standards, multi-actor coordination with weak collaboration and low application of technology [8].

Based on these issues, this study uses literature review and inductive analysis to examine the financial data governance framework and data quality evaluation in the big data environment. The paper first analyzes the practical difficulties of financial data governance, then discusses the main stages and functions of the governance framework. It further summarizes the data quality evaluation indicator system and compares the applicable scenarios and limitations of common evaluation methods. By clarifying the relationship among governance frameworks, quality evaluation, and feedback improvement, this study aims to offer ideas for improving financial data governance in financial institutions.

2. Related Studies and Practical Challenges

Previous research on financial data governance and data quality assessment in the big data context has primarily centered on regulatory data governance, risk avoidance, and framework of digital finance governance. While these studies focus on varying aspects, there is one common thread: Financial data governance is not just a data management problem. It is tightly coupled with regulations, risk management and business processes.

As for regulatory governance, Li Jia pointed out that financial data has sensitive features, it moves easily and is prone to risk spillovers. Existing challenges encompass lack of clarity on who owns the data, challenges in cross-institutional data flow, delayed regulations, and ambiguities regarding the security responsibilities. This view points to a need for financial data governance that is not (solely) dependent on internal technical management in financial institutions. It must also facilitate data sharing, privacy measures, proper usage, and risk management within the regulatory landscape.

How the data governance impacts risk identification is more discussed in the research on risk prevention. Shang Bowen believes that the governance ability of financial data directly affects the risk identification, monitoring and early warning effect [10]. Risk signals can be diluted or even misunderstood if the data that underlie them are missing, delayed, duplicated, or not uniformly defined, or if they are hard to integrate across systems. From this point of view, financial data governance is not only connected with the ability to use data resources, but also with the ability to discover and handle financial risks in time.

Liu Guoping suggested in the research of digital finance governance system, that the systematic governance should be established based on data resource management, data standardization, data quality control, data security protection, and data application services [11]. This perspective suggests that financial data governance ought to extend beyond just data cleansing and technical processing to encompass how data is collected, stored, processed, shared, applied, and fed back.

Existing studies and actual practice of financial institutions demonstrate that governance issues are seldom determined with a single stage. Different systems put data into different formats, causing data fragmentation and separation of systems, such as credit systems, payment systems, risk control systems, auditing systems, customer management systems, and regulatory reporting systems. For identical data, data systems may assign different names to fields, different code to the data, different definitions of statistics, etc., making integration and subsequent analysis more challenging. Modelling analysis, identification of risks and regulatory reporting can also be impacted by missing, duplicate, inaccurate, late or illogical data. Meanwhile, financial data are sensitive, including

security of funds and customer privacy, meaning that data usage must comply with security and regulatory needs as well.

Existing research has provided useful theoretical foundations and practical insights for financial data governance, but further integration is still needed. Some studies focus more on macro-level regulation or risk governance, while giving limited attention to how internal stages such as data access, standard unification, quality control, and application feedback are connected within financial institutions. Other studies discuss data quality evaluation indicators and methods, but do not fully explain how evaluation results can feed back into the governance process. Based on this gap, the following sections analyze the governance framework, data quality evaluation indicators, and evaluation methods to further clarify the structure and evaluation path of the financial data governance system.

3. Construction of the Financial Data Governance Framework

3.1 General Logic of the Financial Data Governance Framework

In the big data environment, financial data governance needs to cover the entire process of data generation, access, processing, storage, use, and feedback. Financial data come from a wide range of sources, including internal business systems such as credit, payment, risk control, auditing, customer management, and regulatory reporting, as well as external market platforms and third-party data channels. Without a unified governance process, different systems may easily lead to fragmented data, inconsistent indicator definitions, unstable data quality, and unclear responsibility boundaries. Therefore, financial data governance should not remain limited to data cleaning or system construction. It needs to form an integrated framework that covers business processes, technical support, management responsibilities, and compliance requirements.

Operational standpoint, financial data governance might adhere to the reasoning of data access, standard unification, centralized integration, quality control, security compliance and application feedback. On the front-end, financial institutions must define the sources of data, how the data are collected and how often, etc., to ensure that data from various sources are fed into the system in a uniform manner. Once data are introduced into the governance process, then field standards, coding rules, indicator definitions, and data dictionaries are required to minimize data understanding disparities across data systems. Then, customer, account, transaction, credit and risk data stored in various business systems can be consolidated with the use of data warehouses, data lakes or master data management systems to break down data silos.

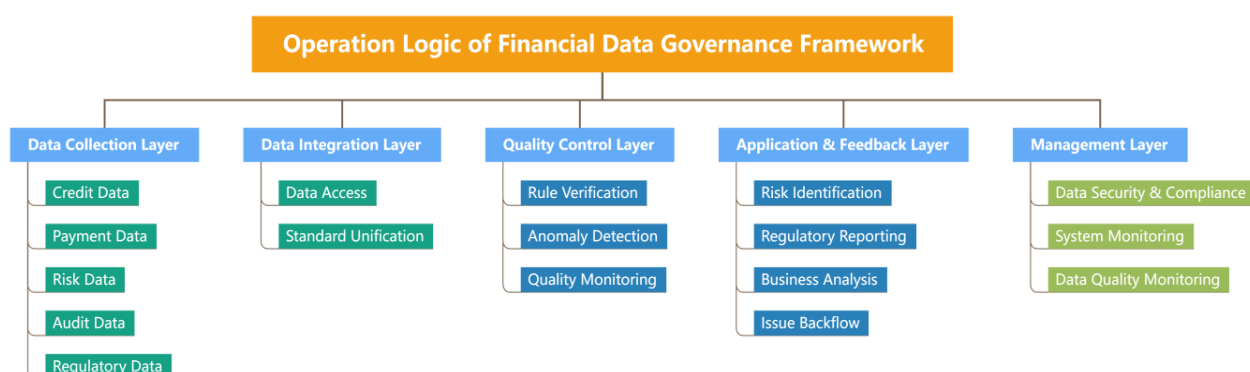


Figure 1: Operational Logic of the Financial Data Governance Framework

Financial institutions should enhance the accuracy of the data post integration by using rule checking, anomaly detection and quality monitoring. Concurrently, security risks should be handled throughout data circulation and use by access control, data desensitization, encrypted transmission, and log auditing. The governed data is then finally used for business situations like risk identification, regulatory reporting, business analysis, customer service and so on. The problems identified in the application, such as missing data, inconsistent definitions, data delay and data access issues should be reverted to earlier phases, such as data access, and standard unification and quality control, to ensure that the governance process is adapted continuously. From this logic, the financial data governance framework can be summarized as a closed loop process as described in figure 1.

3.2 Core Components of the Financial Data Governance Framework

The financial data governance framework can be divided into six main components: data access, standard unification, storage integration, quality control, security compliance, and application feedback. These components correspond to data entry, standardized expression, centralized management, quality improvement, security constraints, and business feedback. Together, they help transform scattered financial data into usable, verifiable, and traceable data assets.

Access to data is the first step in financial data governance. Typically, financial institutions have to get information from many different sources, such as internal business systems producing customer, account, transaction, credit and risk control information, credit information, market data and regulatory reporting data. When access rules are not unified, there are potential issues with unclear data sources, incomplete data collection, and inconsistent interfaces. Financial institutions thus must establish data sources, interface standards, update cycles, and boundaries of usage; in order to ensure that data that flows into the data governance system is reliable and compliant with basic requirements.

Standard unification is to solve the issue of understanding mismatch between systems. In the long run of operation, various business departments can have their own field names, data formats, coding rules and definition of indicators. Consequently, the same customer, account, transaction or risk indicator can be represented in various ways across different systems. One key requirement for financial institutions is to describe core data in a consistent manner using data dictionaries, field specifications, and indicator definitions to minimize deviations in subsequent integration, analysis and reporting.

Storage integration is about disparate data and siloed systems. In many financial institutions, there are several concurrent business systems, and customer data and transaction data, credit data and risk data are located on separate systems. This makes it more difficult to use data from other departments and to integrate the data for cross-department analysis. You can centralize all these data sources and create a relatively cohesive view of data with data warehouses, data lakes or master data management platforms. This offers data quality assessment, risk detection, and business analysis capabilities.

The quality control is used to decide whether or not the data can be relied on. Data gaps, redundant data, inconsistencies, outdated information, and logical issues can impact credit approval, risk assessment, business intelligence, and regulatory reporting in the financial sector. By implementing rule checking, anomaly detection, duplicate detection, quality reports and continual monitoring, financial institutions can detect and address data quality issues which will minimize the influence of poor data quality on business decision making.

Security compliance is about risk parameters of data moving and using. Financial data includes customer identity, account data, transaction data and fund security. Non-compliance risks and

reputation losses can occur if there is data leakage, unauthorized access, or the lack of privacy protection or improper sharing. To safeguard data security during data collection, transmission, storage and use, financial institutions should integrate data classification, authorization to access the data, sensitive information masking, data encryption, operation logs and compliance review.

Feedback from applications points to the relationship of governance outcomes and business scenarios. Governed data are then utilized for risk control, audit supervision, regulatory reporting, identification of customers, and business choice making. If problems are discovered in the use of the business, for example a missing field, different definitions of a field, delay in the data or data access is hard, it should be reported back to the earlier stages of data access, standard unification and quality control, so that the governance mechanism can be modified over time.

Financial data governance is not a technical process carried out in a single stage. It is a management process in which multiple components are connected with one another. Only when data access, standard unification, storage integration, quality control, security compliance, and application feedback work together can the reliability and business applicability of financial data be improved. To present the governance tasks and functions of each component more clearly, the core components of the financial data governance framework are summarized in Table 1.

Table 1 Core Components and Functions of the Financial Data Governance Framework

Governance Stage	Main Task	Problem Addressed	Expected Output
Data Access	Clarify data sources, interface standards, and update frequency	Complex data sources and incomplete collection	Standardized data entry
Standard Unification	Establish field standards, coding rules, and data dictionaries	Inconsistent field names and statistical definitions	Unified data expression
Storage Integration	Build data warehouses, data lakes, or master data platforms	Data silos and fragmented systems	Integrated data view
Quality Control	Check missing values, duplicate records, abnormal values, and logical conflicts	Data errors, delays, and duplication	Reliable data resources
Security Compliance	Implement access control, desensitization, encryption, and log auditing	Data leakage and unauthorized access	Secure and controllable data environment
Application Feedback	Feed problems back to earlier governance stages	Separation between governance and business application	Continuous optimization loop

4. Financial Data Quality Evaluation System

4.1 Financial Data Quality Evaluation Indicators

The purpose of financial data quality evaluation is to determine whether governed data can effectively support business operations, risk identification, regulatory reporting, and management decision-making. For financial institutions, the value of data should not be measured only by its scale or storage volume. It also depends on whether the data are authentic, complete, stable, secure, and traceable in terms of source and use. Based on the characteristics of financial business, this study evaluates financial data quality from seven dimensions: completeness, accuracy, consistency, timeliness, uniqueness, security, and traceability.

In practical business scenarios, completeness is first reflected in whether key fields are available.

Missing customer identity information, transaction time, loan amount, or repayment status may directly affect customer identification, credit approval, and risk assessment. Accuracy concerns whether the data are consistent with actual business facts, such as whether the transaction amount is correct, whether the account status has been updated, and whether customer information and risk levels match the real situation. Consistency mainly addresses differences across systems. For example, the same customer's identity information should remain consistent in the customer management system, credit system, and risk control system. If different systems use inconsistent definitions or records, later business analysis and regulatory reporting may be affected.

Timeliness refers to whether data collection, updating, and transmission can meet the response requirements of business scenarios. In fraud detection, transaction monitoring, and risk warning, delayed transaction data may cause the system to miss the best time for intervention. Uniqueness concerns whether the same customer, account, contract, or transaction appears repeatedly in the data. Duplicate records may distort statistical results and affect risk judgment. Security requires effective protection during data storage, transmission, and use, especially through access control, data masking, encryption mechanisms, and abnormal access management. Traceability emphasizes whether the data source, processing process, modification records, and usage path are clear. This dimension is particularly important in regulatory reporting, audit inspection, and risk responsibility tracking.

Different financial businesses place different levels of emphasis on data quality indicators. Credit business relies more on the completeness and accuracy of customer information. Risk identification requires timely data updates and consistent definitions across systems. Regulatory reporting depends more on traceability and compliance. Based on the above analysis, the financial data quality evaluation indicator system is summarized in Table 2.

Table 2 Financial Data Quality Evaluation Indicator System

Evaluation Dimension	Meaning	Typical Problem	Application Scenario
Completeness	Whether records and key fields are complete	Missing customer identity information or transaction fields	Credit approval and customer identification
Accuracy	Whether data reflect real business facts	Incorrect transaction amount or account status	Risk assessment and business processing
Consistency	Whether data expressions are unified across systems	Inconsistent customer information across platforms	Regulatory reporting and business analysis
Timeliness	Whether data collection and updating meet business requirements	Delayed transaction data or outdated risk data	Anti-fraud and risk warning
Uniqueness	Whether duplicate customers, accounts, contracts, or transactions exist	Duplicate customer records or repeated transaction entries	Customer management and statistical analysis
Security	Whether data are effectively protected during storage, transmission, and use	Unauthorized access or sensitive information leakage	Data sharing and compliance management
Traceability	Whether data sources, processing steps, and usage paths are clear	Missing modification records or unclear responsibility tracking	Audit inspection and regulatory reporting

4.2 Financial Data Quality Evaluation Methods

Financial data quality evaluation should not only identify problems in the data itself, but also allow the evaluation results to be used in later governance activities, such as adjusting data standards, improving quality rules, and refining responsibility mechanisms. Based on the practical needs of financial data governance, this study classifies data quality evaluation methods into three types: rule checking, indicator-based evaluation, and comprehensive evaluation with feedback.

Rule checking is a basic method for data quality inspection. It usually relies on business rules and data specifications to detect missing fields, format errors, abnormal values, logical conflicts, and duplicate records. For example, transaction amounts should not be negative, loan maturity dates should not be earlier than loan disbursement dates, and customer identity fields should meet basic format requirements. This method is suitable for routine monitoring and basic error detection. It is relatively easy to apply, and the results are straightforward. However, rule checking depends heavily on predefined conditions. Its ability to identify cross-system association problems or hidden data anomalies is still limited.

Indicator-based evaluation sets evaluation criteria around dimensions such as completeness, accuracy, consistency, timeliness, uniqueness, security, and traceability, and then uses these criteria to quantify or classify financial data quality. To compare different systems, business lines, or data topics, a comprehensive quality score can be constructed. The basic formula is as follows:

$$DQ=w_1S_1+w_2S_2+w_3S_3+\cdots+w_nS_n$$

In this formula, DQ denotes the comprehensive data quality score; w_i denotes the weight of the i -th evaluation indicator; S_i denotes the score of the i -th evaluation indicator; and n denotes the number of indicators. The weights should not be set independently of business context. They may be determined according to business importance, expert judgment, or actual application needs. For example, traceability and consistency are usually more important in regulatory reporting, while timeliness and accuracy often require higher weights in fraud detection.

Indicator-based evaluation can produce relatively clear data quality results and help financial institutions identify weak points across different systems, business lines, or data topics. In practice, however, indicator weights and scoring standards may be influenced by business experience, management objectives, and the evaluation subject. Without scenario-based adjustment, the evaluation may remain at the level of scores and provide limited support for governance improvement.

The emphasis of comprehensive evaluation with feedback is on the possibilities for using the outcomes of the evaluations to provide feedback to the governance process. Financial data quality evaluation shouldn't end with data problem identification or data quality scoring. It should also direct the access to information, standard unification, storage integration, quality control, and security compliance. For instance, if it's the same fields that are missing, it could mean front-end collection rules need tweaking. Data dictionaries and indicator definitions should be reviewed as there is persistency of indicator definitions within systems. If there is a lot of unusual access to sensitive data, the level of access control and log auditing may need to be increased.

When the evaluation indicators are clear, if there is enough sample data and further quantitative comparison is needed in the scenario, the analytic hierarchy process, entropy weight method or fuzzy comprehensive evaluation can be introduced in the evaluation. These methods can be used to find indicator weights, or to solve multidimensional evaluation problems. They should, however, be applied in the context of specific business scenarios and not as stand-alone tools. Financial data quality assessment is not about getting to the n th degree of sophistication but about finding out where the data issues are and encouraging data governance.

Different evaluation methods serve different governance stages and objectives. Rule checking is more suitable for basic monitoring, indicator-based evaluation is useful for comprehensive judgment, and comprehensive evaluation with feedback helps support a governance loop. In practice, financial institutions should choose appropriate methods according to data types, business scenarios, and evaluation objectives, rather than reducing data quality evaluation to a simple scoring exercise.

5. Conclusion

In the big data environment, financial institutions increasingly rely on high-quality data for risk identification, regulatory reporting, business analysis, and decision-making. This study reviewed the financial data governance framework and data quality evaluation, and summarized the practical challenges, framework components, evaluation indicators, and evaluation methods involved in financial data governance. The analysis suggests that the complexity of financial data governance mainly comes from diverse data sources, fragmented systems, inconsistent standards, unstable data quality, and strict security and compliance requirements. Missing, duplicated, delayed, incorrect, or inconsistent data across systems may directly affect key business activities such as credit approval, risk warning, fraud detection, and regulatory reporting.

Financial data governance from a governance framework standpoint includes access to financial information, standardization, storage integration, information quality control, security compliance and application feedback. It is not meant to be a mere introduction of a certain technical tool, but to solve issues like data isolation, inconsistent data definitions, data quality fluctuations, security risks etc. via institutional framework, process coordination, responsibility assignment and business scenarios. The standardization, reliability and controllability of financial data can only be improved when a relatively stable governance mechanism is accomplished for data collection, circulation, processing and application.

Data quality evaluation for financial data should be done in a multi-faceted way, comprising the following aspects: completeness, accuracy, consistency, timeliness, uniqueness, security and traceability. Data quality requirements vary from business to business. The more thorough and precise customer information is in a credit business, the better. The identification of risks needs to be done in real time with data updates and standardized definitions across systems. Traceability and compliance are more important with regulatory reporting. So, data quality evaluation shouldn't just be single-indicator checks or the results of scores. It should connect to business scenarios, data types, and governance goals and it should lend itself to problem tracing, cause analysis, and subsequent improvement.

Prior studies have focused on the standardization of financial data, financial data quality control, security compliance and development of digital finance. But, the relationship between governance structures and quality assessment remains to be closer. When evaluating quality in business practice, the evaluation should not just result in a score or report. It should be integrated back into certain governance activities such as the data sources, field definitions, system interfaces, owning departments, and security management. Future financial data governance efforts should focus more on process monitoring and business feedback to identify and resolve data issues at the point of data collection, processing, usage, and reporting. Financial data governance is the blend of institutional structures, tools, and business mandates that can help prevent risks, meet regulatory requirements, and drive financial digital transformation.

References

- [1] Wang, W. M., & Yin, S. R. Making full use of quantitative analysis technology to improve the utilization rate of financial data [J]. *Southwest Finance*, 2002(12): 2.
- [2] Wu, X. D. Financial data-based auditing from the perspective of big data [N]. *Accounting Information News*, 2025-01-06(007).
- [3] Cao, P., & Xu, J. P. The logic, mechanism, and path of national audit in optimizing financial data governance [J]. *Finance and Accounting Research*, 2025(9): 52-59.
- [4] Gao, H. S. Strengthening data quality governance and building a solid “digital foundation” for financial innovation [J]. *China Financial Computer*, 2022(6).
- [5] Zhang, Z. (2026). Research on Performance Monitoring and Data-driven Optimization Mechanisms for AI Systems Oriented Toward Decision Support. *Advances in Computer and Communication*, 7(2).
- [6] Wu, Y. (2026). Research on Interpretable Confidence Rule Base Modeling Method Integrating Data-Driven and Constrained K-Means Optimization. *Procedia Computer Science*, 279, 612-619.
- [7] Gao, Y. (2026). Governance Structures and Checks and Balances in Private Equity Funds: Practice, Problems, and Improvement Paths. *Financial Economics Insights*, 3(2), 82-91.
- [8] Liu, Z. (2026). Research on Measuring User Behavior Response Differences Supported by Propensity Scoring Method. *European Journal of AI, Computing & Informatics*, 2(2), 47-53.
- [9] Wu, Y. (2026). Research on Interpretable Confidence Rule Base Modeling Method Integrating Data-Driven and Constrained K-Means Optimization. *Procedia Computer Science*, 279, 612-619.
- [10] Sun, J. (2026). Automated Feature Engineering and Screening System for Large-Scale Factor Libraries. *Procedia Computer Science*, 281, 1282-1290.
- [11] Hou, Y. (2026). Collaborative Regulation for Stable Data Center Operation under Energy Efficiency Constraints. *Procedia Computer Science*, 281, 612-620.
- [12] Zhang, Z. (2026). Research on Performance Optimization Methods for Resource-Aware Model Services in AI Systems. *Procedia Computer Science*, 281, 1310-1317.
- [13] Chang, C. W. (2026). Privacy Infrastructure Vulnerability Mining and Automated Framework Based on Multimodal AI. *Procedia Computer Science*, 279, 702-711.
- [14] Wu, L. (2026). Construction and Evolutionary Analysis of a Game Model for Supply Chain Finance Funding Based on Blockchain Technology. *Procedia Computer Science*, 282, 2004-2012.