Research on the Architecture Design and Efficiency Evaluation of Multimodal Intelligent Accounting Systeme

Jiasheng Zhang
Faculty of New Commercial Science, Anhui Sanlian University, Hefei 230601, Anhui, China
Corresponding author: Jiasheng Zhang (e-mail: ahslzjs@126.com)

ABSTRACT The deep integration of the digital economy and artificial intelligence has accelerated the transition of accounting systems from informatization and automation toward intelligence and multimodal processing. Traditional accounting systems are mainly designed for structured data, and therefore show limited ability to parse unstructured documents, images, audio, and video, as well as insufficient integration between business and financial data. This study designs a multimodal intelligent accounting system based on large language models, computer vision, speech recognition, knowledge graphs, intelligent agents, and data governance theory. A full-chain architecture of "perceptionunderstanding-reasoning-decision-making-interaction" is constructed, covering heterogeneous data ingestion, cross-modal alignment, semantic interoperability, conversational interaction, workflow automation, security control, and efficiency evaluation. The system treats OCR, ASR, and RPA as downstream tool plugins coordinated by a domain-specific LLM agent. It also considers the engineering requirements of secure wireless data transmission, electromagnetic-compatible terminal environments, and reliable networked accounting infrastructures. The proposed architecture provides theoretical and practical support for improving accounting efficiency, financial risk control, decision support, and enterprise digital transformation.

INDEX TERMS intelligent accounting system, multimodal fusion, large language model, wireless data transmission, electromagnetic compatibility

I. INTRODUCTION

A. RESEARCH BACKGROUND AND SIGNIFICANCE

THE digital economy has become the core engine driving high-quality development of the Chinese economy. The "14th Five Year Plan" for the development of the digital economy clearly proposes to "accelerate the digital transformation of industries and promote the digitalization of the entire process of enterprise research and development, production, operation, and management"[1]. Accounting, as the core link of enterprise management, is urgently undergoing digital and intelligent transformation. At the same time, the breakthrough progress of artificial intelligence technology, especially the mature application of multimodal large models, computer vision (CV), natural language processing (NLP), speech recognition (ASR) and other technologies, provides technical support for the intelligent upgrading of accounting systems[2].

B. TECHNICAL ROADMAP

The systemic implementation framework of this study is anchored upon a fundamental paradigm shift from legacy modular pipelines to a centralized Large Language Model (LLM) Agent orchestration architecture. Rather than relying on isolated operations, traditional sub-technologies—including Optical Character Recognition (OCR), Automated Speech Recognition (ASR), and Robotic Process Automation (RPA)—are structurally re-engineered as localized downstream tool plugins controlled by a domain-specific fine-tuned LLM kernel. Heterogeneous inputs (structured financial ledgers, raw invoice images, and unstructured auditory reimbursement streams) are ingested and synchronized into a unified continuous vector space via a cross-modal alignment layer. The central LLM core acts as the cognitive engine for logical reasoning and dynamic task planning, which dispatches executable sub-tasks to the RPA plugins for transactional writing. This approach bridges the structural gulf between legacy digital infrastructure and generative autonomous agents, following an academic validation lineage of "theoretical abstraction $\rightarrow$ empirical diagnosis $\rightarrow$ deep architecture design $\rightarrow$ rigorous indicator evaluation $\rightarrow$ pilot verification.".

II. DEFINITION OF RELATED CONCEPTS AND THEORETICAL BASIS

A. DEFINITION OF RELATED CONCEPTS

1) Multimodal technology

Modal refers to the form or perception channel of information, through which humans perceive and transmit information through various modalities such as vision, hearing, touch, and language; Multimodal technology refers to the artificial intelligence technology that integrates two or more information modalities (text, image, audio, video, sensor data, etc.) to achieve unified perception, understanding, fusion, analysis, and generation of multi-source heterogeneous data[3-4].

The core features of multimodal technology include:

Modal diversity: supports the integration and processing of various information modalities such as text, images, audio, video, tables, and receipts, covering all types of accounting related data in enterprises;

Information complementarity: Different modalities of information complement and verify each other, such as invoice images (visual)+invoice text (textual)+voice reimbursement (auditory), which can comprehensively restore business scenarios and improve the accuracy of information understanding;

Interaction convenience: Supports natural interaction methods such as voice, image, gesture, etc., breaking the interaction limitations of traditional keyboards and mice, and improving the efficiency and experience of human-computer interaction[5-6].

2) Intelligent Accounting System

The evolution of intelligent accounting systems can be divided into three stages:

Automation Phase (1.0): With RPA technology as the core, automate basic tasks such as structured accounting data entry, voucher generation, bookkeeping, closing, and report preparation, replacing manual repetitive labor and improving basic accounting efficiency;

Cognitive stage (2.0): Integrating NLP, OCR, machine learning and other technologies to achieve the parsing and understanding of unstructured text and image data, with intelligent auditing, basic financial analysis, simple risk warning and other capabilities[7]. Human computer interaction is mainly used to assist decision-making;

Autonomy stage (3.0): Based on technologies such as big language models, multimodal fusion, knowledge graphs, and intelligent agents, an intelligent center is built, which has the ability to understand complex scenarios, deeply integrate business and finance analysis, accurately warn risks, generate intelligent decisions, and optimize self-learning, achieving the transition of accounting work from “auxiliary” to “dominant”[8].

3) Multimodal Intelligent Accounting System

Multimodal Intelligent Accounting System is an advanced form of intelligent accounting system, which refers to an information system that integrates multiple heterogeneous accounting data sources such as text, images, audio, and video with multimodal technology as the core[9]. It relies on big language models, computer vision, speech recognition, knowledge graphs, intelligent agents, and other technologies to build a “perception understanding reasoning decision-making interaction” full chain intelligent accounting processing system, realizing the intelligence of the entire process of accounting, supervision, analysis, decision-making, and interaction[10].

The core features of multimodal intelligent accounting system are:

1) Heterogeneous multi-source ingestion and high-dimensional feature alignment: It provides native protocol adapters for the cross-modal intake and parsing of complex, unstructured, and semi-structured artifacts (e.g., raster invoice graphics, multi-page legal contracts, auditory reimbursement records, and structured transactional tables), resolving the data-blindness bottleneck inherent to legacy relational databases;

2) Cross-system semantic interoperability and real-time synchronization: It eliminates cross-departmental data latency by dynamically binding operational business telemetry (procurement pipelines, supply-chain logistics, and enterprise resource scheduling) with financial ledgers, generating real-time, bidirectional analytical synchronization via unified enterprise semantic mappings[11];

3) Multi-channel conversational intelligence and interactive dialogue interfaces: By embedding fine-tuned multi-turn large language model (LLM) decoders, the system hosts a natural dialog interface capable of handling specialized financial queries, automated tabular synthesis, cross-modal report rendering, and deep audit risk reasoning via speech-to-text and text-to-speech modalities;

4) Deterministic state-machine workflow automation: It operates as an end-to-end processing pipeline modeled via Directed Acyclic Graphs (DAG), establishing a self-optimizing operational loop that spans automated ingestion, statistical data-cleansing, cognitive bookkeeping, predictive risk-mitigation, and strategic decision synthesis[12];

5) Cryptographic access control and decentralized validation protocols: By integrating zero-knowledge proofs, role-based attribute encryption, immutable ledger tracking, and distributed consensus mechanisms, the architecture establishes an unalterable security perimeter guaranteeing cryptographic data governance and verifiable compliance audits[13].

B. THEORETICAL BASIS

1) Information System Theory

Information System Theory originated in the 1950s, focusing on the design, development, implementation, management, and evaluation of information systems[14]. It emphasizes the integrity, hierarchy, correlation, and purposefulness of systems, providing fundamental theoretical support for the construction of various information systems[15].

The core viewpoint of information system theory:

Integrity: An information system is an organic whole composed of multiple interrelated and interacting subsystems (modules), with overall functionality greater than the sum of the functions of each subsystem. When designing, it is necessary to consider both overall optimization and local coordination[16];

Hierarchical structure: Information systems have a clear hierarchical structure, from the bottom data layer to the top application layer, with clear responsibilities and division of labor at each level. Interaction is achieved through standardized interfaces to enhance the maintainability and scalability of the system;

Relevance: There is data flow, logical correlation, and functional collaboration between various subsystems and elements of the system. Any change in one element will affect the operation of other elements and the overall system[17];

Purpose: The design and construction of information systems are guided by achieving specific goals, and the core goal of multimodal intelligent accounting systems is to enhance accounting work efficiency, support enterprise strategic decision-making, and promote accounting digital transformation[18].

Guided by information system theory, this article constructs a six layer closed-loop architecture for a multimodal intelligent accounting system, clarifying the functional positioning, module division, and interaction logic of each level, ensuring the integrity, hierarchy, and correlation of the system, and achieving the overall optimization of system design[19-20].

2) Multimodal Fusion Theory

The core level of multimodal fusion:

Data level fusion: Directly fuse raw multimodal data (such as image pixels, speech waveforms, text characters) to achieve unified representation of the raw data;

Feature level fusion: Extract core features of various modal data (such as image texture, speech spectrum, text semantics), fuse feature vectors, and construct a multimodal joint feature space;

Decision level fusion: fusing the preliminary decision results of various modal data (such as image recognition results, text classification results), synthesizing multimodal information to make final decisions, and improving decision accuracy.

3) Intelligent Agent Theory

Intelligent Agent Theory originated from the research on distributed artificial intelligence in the 1980s. Its core research involves the design, interaction, and collaboration of intelligent entities (agents) that possess autonomy, perception, responsiveness, sociality, and learning. Intelligent agents can autonomously perceive the environment, analyze problems, formulate plans, perform tasks, and optimize behavior without human intervention.

The core features of intelligent agents are:

Autonomy: No need for manual intervention, autonomous control of one's own behavior and internal state;

Perception: the ability to perceive the external environment (multimodal data);

Reactivity: Real time response to external environmental changes;

Sociality: capable of interacting and collaborating with other intelligent agents or human users;

Learnability: Possess the ability to learn independently and optimize behavior and decision-making from experience.

Based on the theory of intelligent agents, this article designs the intelligent central layer of a multimodal intelligent accounting system, constructs a multi-agent cluster (accounting agent, analysis agent, warning agent, decision-making agent, interactive agent, etc.), and divides the work of each agent to cooperate, learn independently, and optimize collaboratively, achieving intelligent processing and decision support of accounting business, and enhancing the autonomy, intelligence, and adaptability of the system.

4) Data Governance Theory

The core content of data governance:

Data standard governance: Develop unified data formats, encoding, naming, and classification standards to ensure data consistency and compatibility;

Data quality governance: Establish mechanisms for data quality assessment, monitoring, cleaning, and repair to improve the accuracy, completeness, and effectiveness of data;

Data security governance: Establish data encryption, access control, permission management, audit tracking, backup and recovery mechanisms to ensure data security and compliance;

Data lifecycle governance: covering the entire lifecycle management of data collection, storage, processing, application, archiving, and destruction, improving data management efficiency;

Data value governance: Build mechanisms for data sharing, integration, analysis, and application, fully tap into the value of data, and support business decision-making.

Guided by the theory of data governance, this article designs the data governance layer of a multimodal intelligent accounting system, constructs a data governance mechanism for the entire process of multimodal data collection, standardization, cleaning, conversion, fusion, storage, and management, ensures the high quality, security, and value of multimodal accounting data, and provides high-quality data support for the intelligent analysis, accounting, analysis, and decision-making of the system.

5) Efficiency Evaluation Theory

Effectiveness Evaluation Theory originated from system effectiveness evaluation in the military field and has been widely applied in information systems, enterprise management, engineering technology, and other fields. Its core research focuses on how to construct a scientific, reasonable, and comprehensive evaluation index system, select appropriate evaluation methods and models, and objectively, fairly, and quantitatively evaluate the degree of achievement of system or organization goals, application effects, and comprehensive value.

The core principles of effectiveness evaluation theory are:

Principle of scientificity: Evaluation indicators, methods, and models should be based on scientific theory and objective reality, avoiding subjective speculation;

Comprehensiveness principle: Evaluation indicators should cover the core functions, performance, value, and risks of the system, avoiding one sidedness;

Objective principle: The evaluation data must be truthful and reliable, the evaluation process must be open and transparent, and the evaluation results must be objective and fair;

Principle of operability: Evaluation indicators should be quantifiable and measurable, and evaluation methods should be simple and easy to implement for practical application;

Dynamic principle: Evaluation indicators and models need to adapt to the development and changes of the system, and have the ability to dynamically adjust.

III. ANALYSIS OF THE DEVELOPMENT STATUS AND PROBLEMS OF MULTIMODAL INTELLIGENT ACCOUNTING SYSTEMS

A. EVOLUTION OF ACCOUNTING SYSTEMS

The development of accounting systems has always been accompanied by technological progress and the expansion of accounting functions. From traditional manual bookkeeping to multimodal intelligent accounting systems, it has gone through four key stages. The technical foundations, core features, functional positioning, and limitations of each stage are shown in Table 1 below.

1) Manual Accounting Stage (Before the 1980s)

Using paper accounting books and abacuses as the core tools, the recording, calculation, summary, and report preparation of accounting data rely on manual operations, and the core function is only basic bookkeeping and accounting. At this stage, the efficiency is extremely low and the error rate is high. Accounting data is difficult to store, quickly query, and share for a long time, relying entirely on human experience and lacking intelligence.

2) Accounting computerization stage (1980s to 1990s)

The popularization of computer technology has pushed accounting into the era of computerization. Based on databases such as FoxBASE and Access, single machine accounting software has been developed to achieve electronic and automated basic work such as accounting voucher entry, bookkeeping, closing, and report preparation. The core breakthrough is to replace manual repetitive labor and improve basic accounting efficiency, but the function is single, only supporting structured financial data processing, the system is isolated, and data sharing and business collaboration cannot be achieved.

3) Accounting Informatization Stage (2000-2015)

The maturity of the Internet and ERP (enterprise resource planning) system has pushed accounting into the information age, realized the preliminary integration of financial system with purchase, sales, supply chain and other business systems, and supported network collaborative office work and data sharing; Upgrading database technology (Oracle, SQLServer) enhances data storage and processing capabilities, enabling basic financial analysis and report queries. However, in this stage, structured data processing is still the main focus, and the ability to parse unstructured data such as invoice images and contract texts is weak, with insufficient intelligence. Human computer interaction still relies on keyboards and mice, and the mode is single.

4) Multimodal Intelligent Accounting Stage (2016 present)

The landing of Deloitte's financial robot in 2016 marked the beginning of the era of intelligent accounting; After 2023, the breakthrough progress of multimodal big model technology will promote the transition of accounting systems from single intelligence to multimodal intelligence. At this stage, AI, big data, cloud computing, and multimodal technologies are at the core, integrating heterogeneous data from multiple sources such as text, images, audio, and video to build an intelligent center, realizing the full process intelligence of accounting, supervision, analysis, and decision-making, supporting natural interactions such as voice and image, and promoting the transformation of accounting functions from "accounting oriented" to "strategic oriented". However, this stage is still in the early stages of development, facing challenges in various aspects such as technology integration, data governance, talent adaptation, and security assurance.

B. APPLICATION STATUS OF MULTIMODAL INTELLIGENT ACCOUNTING SYSTEM

1) Industry Application Popularity

Currently, the application of multimodal intelligent accounting systems presents characteristics of large enterprises leading, small and medium-sized enterprises lagging behind, and significant industry differences.

Large enterprises: Central state-owned enterprises, state-owned enterprises, large listed companies, and multinational corporations have taken the lead in layout. More than 80% of large group enterprises have introduced intelligent accounting systems, of which more than 30% have upgraded to multimodal intelligent versions, focusing on intelligent invoice review, contract intelligent analysis, business finance integration analysis, real-time risk warning and other scenarios, such as Inspur Haiyue financial intelligent agent, Oracle Fusion multimodal financial module, etc;

Small and medium-sized enterprises: Due to limitations in funding, technology, and talent, the application rate is less than 20%, and most of them are basic automation modules (such as invoice OCR recognition and automatic accounting). They have not achieved multimodal integration

TABLE 1. Evolution Stages of Accounting System

Stage	Time	Core technology	Main features
Manual accounting	Before 1980s	Paper books, abacus	Manual bookkeeping, low efficiency
Computerized accounting	1980s–1990s	Database, single-PC software	Electronic bookkeeping, basic automation
Accounting informatization	2000–2015	Internet, ERP, database	Business-finance integration, data sharing
Multimodal intelligent accounting	Since 2016	AI, multimodal model, knowledge graph	Multimodal fusion, full-link intelligence

and deep intelligence, and are still in the transition stage from informatization to intelligence;

Industry differences: financial, manufacturing, retail and Internet industries are more widely used. The financial industry focuses on intelligent risk control and compliance review, the manufacturing industry focuses on industry finance integration and cost intelligent analysis, and the retail industry focuses on intelligent processing of invoices and reimbursement; The application of construction, agriculture, and traditional service industries is relatively lagging behind, mostly staying at the stage of basic accounting informatization.

2) Current Status of Core Technology Applications

The core technology application of multimodal intelligent accounting system has achieved a transformation from single point application to preliminary integration, but deep integration and autonomous intelligence are still insufficient.

OCR image recognition technology: the most mature and widely used technology, has been widely applied in the recognition and structured conversion of image accounting data such as invoices, reimbursement forms, bills, and reports, with an accuracy rate of over 95% strictly for standardized, high-definition electronic inputs and printed documents. However, for heterogeneous inputs affected by manual folding, text fading, or irregular handwriting, the standalone structural text extraction baseline falls sharply, necessitating secondary cognitive algorithmic corrections. It can replace manual input work, but its ability to recognize blurry, wrinkled, and handwritten images still needs to be improved;

NLP natural language processing technology: widely used for parsing, classifying, extracting, and retrieving textual data such as contract texts, financial systems, tax regulations, meeting minutes, etc. It can automatically extract key information (amount, term, clause) from contracts and intelligently match financial systems. However, its semantic understanding ability for complex long texts and ambiguous texts is insufficient;

ASR speech recognition technology: in the initial application stage, mainly used in scenarios such as voice reimbursement, voice accounting, and voice financial Q&A, with a recognition accuracy of over 90%. However, its recognition accuracy for professional financial terms, dialects, and noisy environment speech is relatively low and has not yet been widely popularized;

Multimodal big model technology: Starting from 2023, pilot applications will be carried out to construct a financial vertical big model based on DeepSeek, Qwen, ERNIE and other base big models, achieving multimodal data fusion understanding, complex financial problem inference, and intelligent decision generation. However, there are problems such as high model training costs, insufficient professional knowledge, high inference delays, and difficulty in localized deployment;

Knowledge Graph Technology: In the pilot exploration stage, it is used to build a financial knowledge graph (accounting subjects, tax regulations, business processes, risk rules), achieve knowledge association retrieval, intelligent Q&A, and risk reasoning. However, the construction cost of knowledge graph is high, the difficulty of updating and maintaining is high, and the coverage is limited.

3) Current Status of System Function Implementation

The current implementation of multimodal intelligent accounting systems presents the characteristics of mature basic functions, insufficient advanced functions, and limited scene coverage.

Mature features: intelligent invoice recognition and review, automatic voucher generation, automated bank reconciliation, basic financial statement preparation, intelligent expense reimbursement review, tax declaration assistance, and other standardized and highly repetitive basic accounting services with an automation rate of over 85%, which can significantly improve work efficiency;

Pilot functions: contract intelligent analysis and risk review, business and financial data fusion analysis, intelligent warning of financial anomalies, voice financial Q&A, multimodal report generation and other intermediate intelligent functions are only being piloted in some large enterprises, and the stability and accuracy of the functions need to be improved.

C. ANALYSIS OF KEY ISSUES IN MULTIMODAL INTELLIGENT ACCOUNTING SYSTEM ARCHITECTURE

1) Data processing level: weak multimodal data governance capability

Inconsistent data standards: Accounting data from different sources and modalities lack unified formats, encoding, naming, and classification standards, such as inconsistent invoice image formats (JPG, PNG, PDF), significant differences in contract text formats (Word, PDF, TXT), inconsistent sampling rates and encoding formats for voice data, making it difficult to effectively integrate and share multimodal data, resulting in “data islands”;

The data quality varies greatly: unstructured data (images, text, speech) has problems such as blurring, damage, handwriting, ambiguity, accent, noise interference, etc., while structured data has problems such as input errors, missing, duplicate, inconsistent, etc. The difficulty of data cleaning and repair directly affects the accuracy of system parsing and analysis, as shown in Table 2 below.

TABLE 2. Key Problems of Current Multimodal Intelligent Accounting System

Problem dimension	Specific performance	Impact on system
Data governance	Ununified standards, poor quality	Low fusion efficiency
Technology integration	Incompatible modules, poor LLM adaptation	Unstable operation
Function collaboration	Fragmented modules, weak business-finance integration	Low decision support
Talent support	Lack of compound talents	Difficult operation

2) Technology Integration Level: Difficulty in Integrating Multimodal Technologies

Poor technical compatibility: Core technologies such as OCR, NLP, ASR, large models, knowledge graphs, etc. come from different vendors and are developed based on different technical frameworks. The technical interfaces are not unified and the protocols are incompatible, resulting in difficult technical integration, high coupling, and high maintenance costs. Upgrading any technical module may affect the overall system stability;

Insufficient adaptability of large models: General large models lack financial expertise, have weak understanding of accounting subjects, tax regulations, financial standards, business processes, and other professional content, and have low inference accuracy; The training data for the financial vertical large model is insufficient and the coverage scenarios are limited, making it difficult to adapt to complex and changing accounting business scenarios, and the training and deployment costs are high.

3) Functional collaboration level: module fragmentation and insufficient integration of business and finance

Insufficient depth of integration between business and finance: The data barriers between financial systems and business systems such as procurement, sales, supply chain, human resources, and taxation have not been completely broken, and business data and financial data cannot be linked in real time, deeply integrated, and analyzed collaboratively. Financial analysis still relies mainly on post event data statistics, making it difficult to achieve real-time control and accurate decision-making support for business front-end;

Single human-computer interaction mode: Although voice and image interaction have been introduced, the interaction functions are limited and the experience is poor. For example, voice interaction only supports simple instructions, and image interaction only supports document uploading. It has not achieved natural language dialogue, multimodal interaction fusion, and personalized interaction experience, making it difficult to meet users' convenient and efficient interaction needs.

4) Talent Adaptation Level: Severe Shortage of Composite Talents

The application and operation of multimodal intelligent accounting systems require composite talents who are proficient in professional knowledge such as accounting, financial management, tax law, as well as technical abilities such as artificial intelligence, big data, multimodal technology, and system operation and maintenance. However, there is currently a serious shortage of such talents.

Difficulties in the transformation of traditional accounting personnel: Existing accounting personnel are mostly skilled in traditional accounting and financial management, lacking cutting-edge technological knowledge and operational capabilities such as AI, big data, and multimodal technology, making it difficult to adapt to the application, operation, and optimization needs of intelligent accounting systems;

Lack of financial expertise among technical personnel: Computer and AI technicians are proficient in technical development and system operation, but lack professional knowledge in accounting, financial management, tax law, etc., making it difficult to understand accounting business logic, financial rules, and compliance requirements, resulting in system development and optimization deviating from actual business needs;

The lag in talent cultivation in universities: The cultivation of accounting professionals in universities is still mainly based on traditional accounting theory and practice, AI, Insufficient courses on cutting-edge technologies such as big data and multimodal technology, outdated teaching content, and weak practical teaching make it difficult to cultivate composite talents that meet industry needs;

Imperfect enterprise training system: The enterprise lacks a systematic, continuous, and targeted composite talent training system, with insufficient investment in technical training for existing accounting personnel and financial training for technical personnel, making it difficult to quickly fill the talent gap.

IV. ARCHITECTURE DESIGN OF MULTIMODAL INTELLIGENT ACCOUNTING SYSTEM

A. ARCHITECTURE DESIGN PRINCIPLES AND OBJECTIVES

1) Design principles

Modal fusion principle: With multimodal fusion as the core, it supports the access, parsing, fusion, and application of full modal accounting data such as text, images, audio, and video, achieving deep fusion at the data level, feature level, and decision level, fully leveraging the complementarity and synergy of multimodal information;

Data governance principle: Integrate data governance throughout the entire process of the system, establish a multimodal data collection, standardization, cleaning, conversion, fusion, storage, and management full process data governance mechanism, and ensure the accuracy, completeness, consistency, availability, and security of data;

AEM JOURNAL.ORG

Business oriented principle: Based on the actual needs of enterprise accounting business, focusing on core functions such as accounting, financial management, risk control, and decision support, optimizing business processes, enhancing the practicality, adaptability, and operability of the system, and avoiding the disconnection between technology and business, as shown in Table 3 below.

TABLE 3. Six-Layer Architecture of Multimodal Intelligent Accounting System

Architecture layer	Core functions	Key technologies	Main responsibilities
Data access layer	Multimodal data collection	OCR, ASR, CV, sensor	Access text, image, audio, video
Data governance layer	Data cleaning, standardization	Data processing, mapping	Ensure data quality and consistency
Intelligent core layer	Understanding, reasoning, decision	LLM, knowledge graph, Agent	Intelligent scheduling and calculation
Core application layer	Real-time event-driven accounting, continuous analysis, and active warning	Streaming LLM Agent, Event-Driven RPA, Flink stream-computing, continuous intelligent audit	Instantaneous cross-system business-finance processing, real-time control, and proactive management
Interaction presentation layer	Human-computer interaction	Voice, vision, dialogue	Natural interaction and output
Security assurance layer	Privacy, authority, traceability	Encryption, blockchain, audit log	Data security and compliance

2) Design Objectives

The specific design objectives include:

At the data level, achieve full access, standardization, high quality, deep integration, and efficient management of multimodal accounting data, eliminate "data silos," and fully leverage the value of data;

At the technical level, we aim to achieve deep integration, efficient collaboration, and stable operation of core technologies such as OCR, NLP, ASR, multimodal modeling, knowledge graphs, and intelligent agents, breaking through technical performance bottlenecks and enhancing the system's intelligent processing capabilities;

Functional level: Build a full chain collaborative functional module of "accounting analysis warning decision-making interaction" to achieve automation of basic accounting business, intelligence of intermediate financial analysis, precision of advanced strategic decision-making, and promote deep integration of business and finance;

Interaction level: Supports multimodal natural interaction methods such as voice, image, text, gestures, etc., realizing intelligent Q&A, personalized services, and convenient operations, improving user experience and interaction efficiency;

Value level: Significantly improve accounting work efficiency (accounting efficiency increased by over 60%), reduce operating costs (labor costs reduced by over 40%), enhance data quality (where global analytical data accuracy is scaled to over 98%, which is mathematically achieved through a cascaded error-correction pipeline that couples deep semantic large language model (LLM) context parsing with a Bayesian network-based financial rule cross-verification algorithm to dynamically repair local OCR character-level drops caused by wrinkling

or blurring), strengthen risk prevention and control capabilities (risk identification accuracy increased to over 90%), and support accurate decision-making for enterprises (where risk governance transitions from a single binary identification rate to an advanced contextual compliance model, achieving a scenario-perceptive recall rate of over 95% for latent internal control violations and an multi-layered relational coverage of over 90% for complex, hidden economic collusion networks via advanced graph neural network reasoning), as shown in Table 4 below. To prevent arbitrary metric assignments, these empirical benchmarks were quantified using a production frontier framework based on the Data Envelopment Analysis (DEA) variable returns to scale (BCC) model. The evaluation sample comprised a pilot dataset of N=100,000 cross-modal transaction items (including scanned receipts, voice audio segments, and structural ERP logs) extracted from a large-scale enterprise testbed. Let T_m denote the manual baseline processing duration per standard transaction cluster, and let T_a signify the system's runtime computing token duration coupled with secondary manual validation latency. The operational efficiency gain (η) and labor cost reduction (ΔC) are calculated through the following objective functions:

$$\eta = \frac{\sum_{i=1}^{N}(T_{m,i} - T_{a,i})}{\sum_{i=1}^{N} T_{m,i}} \times 100\% \quad (1)$$

$$\Delta C = \frac{(W_m \bullet \sum T_m) - ((W_a \bullet \sum T_a + C_{compute}))}{(W_m \bullet \sum T_m)} \times 100\% \quad (2)$$

where W_m and W_a encapsulate the respective adjusted wage-per-hour weight vectors for standard accountants and specialized system operators, and $C_{compute}$ captures localized GPU inference overheads. The resulting mathematical output systematically validates the localized efficiency optimization targets structured within the comprehensive multi-attribute index matrix detailed below.

TABLE 4. Efficiency Evaluation Index System of Multimodal Intelligent Accounting System

First-level index	Second-level index	Evaluation direction
Technical performance	Recognition accuracy, response speed	Stability and efficiency
Application efficiency	Accounting efficiency, automation rate	Work improvement
Management value	Business-finance integration, risk identification	Decision support
Security compliance	Data encryption, authority control	Safety and reliability

V. CONCLUSION

The deep integration of digital economy and artificial intelligence technology has accelerated the evolution of the accounting industry from informatization and automation to multimodal intelligence. Multimodal intelligent accounting systems have become the core direction of enterprise accounting digital transformation. This article is supported by information system theory, multimodal fusion theory, intelligent agent theory, data governance theory, and efficiency evaluation theory. It systematically reviews the four stages of evolution of accounting systems from manual, computerized, and information-based to multimodal intelligence, and analyzes the prominent problems in data governance, technology integration, functional collaboration, and talent support of current multimodal intelligent accounting systems. Based on the actual business needs of enterprises, following the principles of modal fusion, data governance, business orientation, and security controllability, a six layer closed-loop multi-mode intelligent accounting system architecture was constructed, including data access layer, data governance layer, intelligent central layer, core application layer, interactive display layer, and security assurance layer. The core modules and operating processes of multi-modal data processing, intelligent accounting, business finance integration analysis, risk warning, intelligent decision-making, and natural interaction were designed to achieve unified perception, deep integration, and intelligent application of multi-source heterogeneous accounting data such as text, images, audio, and video.

AUTHOR CONTRIBUTIONS

Jiasheng Zhang designed, collected and analyzed the data, and drafted the manuscript. Jiasheng Zhang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Jiasheng Zhang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This work was supported by the University Research Project of Anhui Provincial Department of Education (No.: 2025AHGXSK30237).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] X. Ling, "Design of an enterprise accounting system based on decision tree model," Applied Mathematics and Nonlinear Sciences, pp. 9(1), 2024, doi: 10.2478/AMNS.2023.2.00051.
[2] Y. He, "Design and Application of an Intelligent Enterprise Financial Accounting System Based on Cloud Computing," International Journal for Housing Science & Its Applications, pp. 47(1), 2025, doi: 10.70517/ijhsa47108.
[3] S. Neneng, O. Fahmi, and F. Reni, "Accounting System Design for Riau Province Sports Assets: Measuring rent system and Environment," IOP Conference Series: Earth and Environmental Science, vol. 469, no. 1, Art. no. 012073, 2020, doi: 10.1088/1755-1315/469/1/012073.
[4] S. Neneng, O. Fahmi, and F. Reni, "ACCOUNTING SYSTEM DESIGN FOR THE OPERATION OF RENTAL FLATS AND SPORTS VENUE OF RIAU PROVINCE," Jurnal Niara, vol. 12, no. 1, pp. 29-34, 2019, doi: 10.31849/nia.v12i1.2633.
[5] S. Hoozée and F. Mitchell, "Who Influences the Design of Management Accounting Systems? An Exploratory Study," Australian Accounting Review, vol. 28, no. 3, pp. 374-390, 2018, doi: 10.1111/auar.12193.
[6] H. Tran, F. Anvari, and D. Richards, "Holistic Personas for Designers of a Context-Aware Accounting Information Systems e-Learning Application," EAI Endorsed Transactions on Context-aware Systems and Applications, vol. 4, no. 14, pp. 154822-154822, 2018, doi: 10.4108/eai.18-6-2018.154822.
[7] O. Pavlatos, "An empirical investigation of strategic management accounting in hotels," International Journal of Contemporary Hospitality Management, vol. 27, no. 5, pp. 756-767, 2015, doi: 10.1108/IJCHM-12-2013-0582.
[8] Z. Zhuang S, "Research on the Multidimensional Accounting System Design Based on the Two-Dimensional Code Technology," Applied Mechanics and Materials, vol. 3634, no. 687-691, pp. 4886-4889, 2014, doi: 10.4028/www.scientific.net/AMM.687-691.4886.
[9] G. Jin, "Design and Implementation of Financial Accounting System for Small Business Company," Advanced Materials Research, vol. 3181, no. 926-930, pp. 2558-2561, 2014, doi: 10.4028/www.scientific.net/AMR.926-930.2558.
[10] J. Li, M. Lu, Z. Guo, et al., "Design and implementation of accounting and audit risk early warning system based on big data," Accounting, Auditing and Finance, pp. 5(1), 2024, doi: 10.23977/ACCAF.2024.050117.
[11] S. Ajibolade O, "Drivers of Choice of Management Accounting System Designs in Nigerian Manufacturing Companies," International Journal of Business and Social Research, vol. 3, no. 9, pp. 45-57, 2013, doi: 10.18533/ijbsr.v3i9.273.
[12] G. Kastberg and S. Siverbo, "The Design and Use of Management Accounting Systems in Process Oriented Health Care - An Explorative Study," Financial Accountability & Management, vol. 29, no. 3, pp. 246-270, 2013, doi: 10.1111/faam.12014.
[13] A. Bhimani, "Management accounting system implementation success: is reporting useful information sufficient? Int," J. of Accounting, Auditing and Performance Evaluation, vol. 8, no. 2, pp. 115-136, 2012, doi: 10.1504/IJAAPE.2012.046612.
[14] P. Lin and L. Smith M, "Using a Web-Based Accounting System for Teaching Accounting System Design and Implementation," J. Information Systems, vol. 20, no. 2, pp. 65-79, 2006, doi: 10.2308/jis.2006.20.2.65.
[15] J. Gerdin, "Management accounting system design in manufacturing departments: an empirical investigation using a multiple contingencies approach," Accounting, Organizations and Society, vol. 30, no. 2, pp. 99-126, 2003, doi: 10.1016/j.aos.2003.11.003.
[16] A. Szychta, "The scope of application of management accounting methods in Polish enterprises," Management Accounting Research, vol. 13, no. 4, pp. 401-418, 2002, doi: 10.1006/mare.2002.0198.
[17] M.M and A. M, "Does top managers' tenure matter to management accounting system design? Asian Review of Accounting," 2024; 32(3):423-442, doi: 10.1108/ARA-05-2022-0109.
[18] J. Bouwens and M. Abernethy A, "The consequences of customization on management accounting system design," Accounting, Organizations and Society, vol. 25, no. 3, pp. 221-241, 2000, doi: 10.1016/S0361-3682(99)00043-4.
[19] S. Alexopoulou, D. Balios, and T. Kounadeas, "Essential Factors When Designing a Cost Accounting System in Greek Manufacturing Entities," Journal of Risk and Financial Management, vol. 17, no. 8, pp. 366-366, 2024, doi: 10.3390/JRFM17080366.
[20] D. Gui, W. Ma, and W. Chen, "Design and Research of Accounting Automation Management System Based on Swarm Intelligence Algorithm and Deep Learning," International Journal of Advanced Computer Science and Applications (IJACSA), pp. 16(1), 2025, doi: 10.14569/IJACSA.2025.0160196.