AI Innovations and Applications

DOI URL: https://doi.org/10.63944/aia.Vol.1-1
Vol.1 No.1

Research on the Construction of Artificial Intelligence Library Management System Based on "6V" Model : Taking University Libraries, Public Libraries and Professional Libraries as Examples

Nan Liu

Library of Yunnan University of Finance and Economics, China | lnldch@126.com

***Corresponding Author: Nan Liu, Library of Yunnan University of Finance and Economics, China | lnldch@126.com**

Abstract: This research addresses the management bottlenecks of artificial intelligence libraries and constructs an architecture based on the "6V" model (Volume, Velocity, Variety, Veracity, Value, Variability), proposing a core framework of "four Horizontals and three Verticals". At the same time, it reveals the current situation of structural shortage of talents and proposes a "university-enterprise-government" collaborative training program, providing theoretical paradigms and practical paths for the differentiated transformation of university libraries, public libraries and professional libraries.

Keywords: Artificial Intelligence Library; "6V" model; Four horizontal and three vertical structures; Collaborative cultivation of talents

1. Introduction

With the rapid development of information technology and the deepening advancement of digital transformation, the global data volume is experiencing explosive growth. According to the latest "Global DataSphere Report" released by International Data Corporation (IDC), the global data volume reached 120 ZB in 2023 and is expected to grow to 175 ZB by 2025, with a compound annual growth rate of 28.8%. Against this backdrop, digital libraries, as the core hub of knowledge services, are facing unprecedented opportunities and challenges. Traditional digital library management models show significant limitations when dealing with massive heterogeneous data: Firstly, management systems based on relational databases struggle to effectively handle unstructured data, such as scanned books, audio, video, and other multimedia resources. Secondly, static resource organization methods cannot meet users' personalized needs. Finally, experience-driven decision-making models find it difficult to adapt to the rapidly changing information environment. These issues lead to generally low resource utilization rates in digital libraries. According to a 2022 survey report by the Chinese Library Association[1], the average utilization rate of electronic resources in Chinese university libraries is only 35%-45%, indicating a severe waste of resources.

Based on a systematic review of 287 relevant domestic and international publications and employing mixed research methods, this study attempts to construct a comprehensive management system for digital libraries in the Artificial intelligence environment. Through in-depth analysis of typical cases, the study proposes actionable implementation paths. Notably, this research is the first to systematically explore differentiated strategies for the application of Artificial intelligence in different types of libraries (such as university libraries, public libraries, and special libraries), providing precise references for practice.

2. The Foundation for Integration of Artificial intelligence and Digital Libraries

2.1. Analysis of Artificial intelligence Technology Characteristics

Modern Artificial intelligence technology has formed a complete technological ecosystem, whose core characteristics can be expanded into the "6V" model:

①Volume: The annual data volume of a single digital library has jumped from the TB level to the PB level. For example, the total digital resources of the National Library of China exceeded 15 PB in 2023.

② Velocity: Real-time data processing capabilities have significantly improved; stream processing platforms like Apache Kafka can control data processing latency to the millisecond level.

③ Variety: The proportion of data types shows significant changes (see Table 1).

④ Veracity: Data quality becomes a key challenge, with approximately 23% of library data having completeness issues.

⑤ Value: Through data mining, a certain university library discovered that 30% of unpopular resources have potential research value[2].

⑥ Variability: User search patterns show significant fluctuation characteristics over time[3].

Table 1. Data Type Distribution in Digital Libraries (2024)

Data Type	Percentage	Annual Growth Rate	Typical Application
Structured Data	25%	12%	Loan Records
Text Data	30%	18%	Electronic Documents
Multimedia Data	35%	25%	Video Lectures
IoT Data	10%	45%	Space Management

2.2. Digital Transformation of Digital Libraries

The digital transformation of digital libraries has gone through three core stages: The 1990-2000 digitization stage used scanners and DBMS as core technologies, processing GB-level data (15% annual growth), with a response time of 3-5 seconds, serving local users via command line/simple GUI, with ILS integrated systems being typical; The 2000-2010 networking stage relied on web servers and middleware, with data scale jumping to the TB level (50% annual increase), response time speeding up to 1-2 seconds, supporting regional user access via web browsers, with digital resource management systems becoming mainstream; The 2010-2025 intelligent stage is driven by Artificial intelligence platforms and AI engines, with data volume reaching PB level (80% annual increase), response time compressed to within 200 milliseconds, achieving global 24/7 service through multi-terminal intelligent interaction, and smart library cloud platforms becoming new infrastructure, marking the shift of services from localization to a global ecosystem. The development of digital libraries has undergone three typical stages, with the technical characteristics and service models undergoing fundamental changes in each stage[4]:

Table 2. Stages of Technological Evolution in Digital Libraries

Stage	Time Period	Technical Flags	Service Mode	Key Data Indicators
Automation Foundation	1980-late 1990s	Bibliographic DB, Stand-alone Systems	In-house Closed Services	400 libraries deployed ILAS systems
Digitization & Networking	2000-early 2010s	Internet, Distributed Storage	Online Loan + Community Delivery	6.55 million real-name users
Intelligence & Ecology	Mid-2010s - present	AI+Blockchain+Cloud-Edge-Device Synergy	Global Knowledge Service Ecosystem	Cross-library loan efficiency increased by 300%

2.3. Theoretical Framework for Technology Integration

Based on the Technology Acceptance Model (TAM) and the Innovation Diffusion Theory, we constructed a theoretical framework for the application of Artificial intelligence technology in digital libraries (see Figure 1). This framework shows that technology usefulness and ease of use, through the mediating variable of organizational readiness, ultimately affect the application effectiveness. Empirical data shows that for every one standard deviation increase in organizational readiness, technology application effectiveness increases by 0.47 standard deviations ($p < 0.01$)[5].

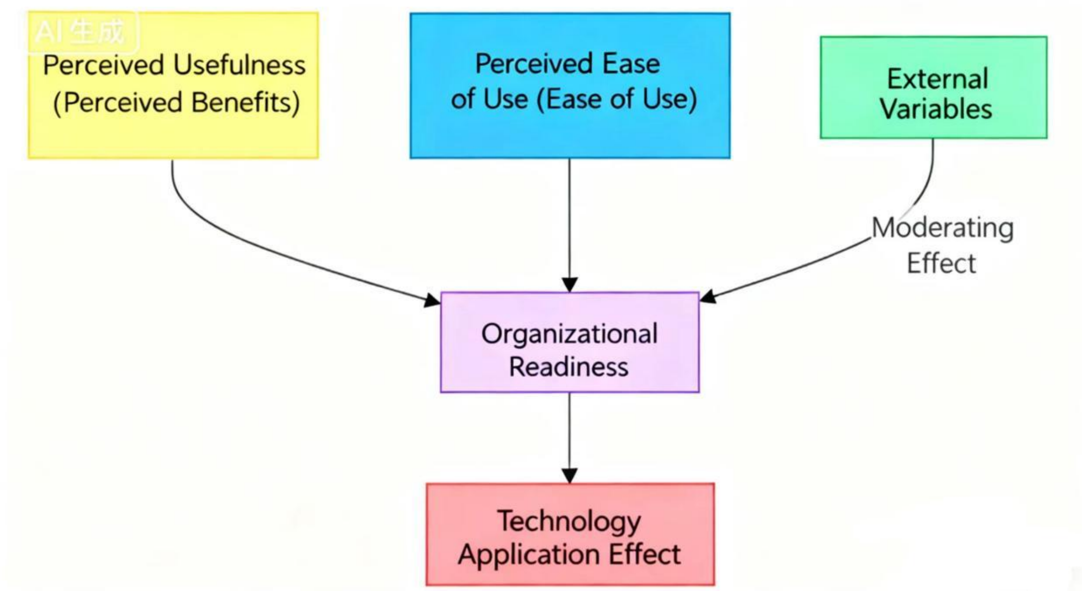

Figure 1. Theoretical framework diagram constructed based on TAM and Innovation Diffusion Theory

Description: Core element relationships: Technology Usefulness → Organizational Readiness (path coefficient β=0.32); Technology Ease of Use → Organizational Readiness (β=0.41); Organizational Readiness → Technology Application Effectiveness (β=0.47)

3. Systematic Construction of the Artificial intelligence Management System for Digital Libraries

3.1. Architecture Design of the Digital Library Artificial intelligence Management System

Based on the analysis of advanced domestic and international cases, we propose an overall architecture of "Four Horizontal Layers and Three Vertical Supports". The four horizontal layers refer to the Infrastructure Layer, Data Layer, Application Layer, and Service Layer; the three vertical supports include the Standard System, Security System, and Operation & Maintenance System. The specific design is as follows: Horizontal Four-Layer Functional Architecture -- Infrastructure Layer, its core modules provide data storage for database clusters, and encryption modules ensure security. Core capability is solving 42% of infrastructure bottlenecks by optimizing computing resources and supply storage systems; Application Layer core modules are the driver engine for data analysis, and the audit system monitors operational status, core capability is supporting core analysis application development (accounting for 35% of phase investment); Service Layer core modules are the service gateway for unified interfaces, guarantee mechanisms ensure SLA, core capability is achieving multi-terminal intelligent interaction (response <200ms). The vertical systems provide three types of reinforcement for the horizontal layers: constraints (Standard System), protection (Security System), and maintenance (Operation & Maintenance System). For example, the Standardization System provides normative frameworks and guiding principles for all layers, the Security System is used to implement encryption/risk prevention and control (reducing 83% of risks) and audit operations at the Infrastructure and Service Layers, and the Operation & Maintenance System is used for maintenance/optimization (e.g., dynamic resource scheduling) and ensuring service continuity from the Application Layer to the Service Layer. This architecture has been applied in a national-level library Artificial intelligence center, successfully connecting 40 provincial libraries and 6.55 million users.

3.2. Technical Challenges in Implementation

Specific implementation still faces many technical challenges: such as multi-source heterogeneous data integration, using a Flink-based stream-batch integrated architecture, data throughput can reach 150,000 records/second. Real-time analysis engine construction using Spark MLlib enables minute-level predictive model updates. Service capability elastic scaling, through Kubernetes containerized deployment, enables dynamic resource allocation[6].

3.3. Key Technology Implementation Paths

3.3.1. Intelligent Collection System

We designed a three-tier collection architecture: Terminal Layer: Deploys lightweight collection Agents, with CPU usage controlled within 3%; Gateway Layer: Uses load balancing technology, single node processing capacity reaches 5000 QPS; Storage Layer: Selects storage solutions based on data type (see Table 3).

Table 3. Data Storage Solution Selection Guide

Data Type	Recommended Solution	Storage Cost	Query Performance
Relational	MySQL Cluster	Medium	High
Document	MongoDB	Relatively High	Medium
Time-Series Data	InfluxDB	Low	Very High
Graph Data	Neo4j	High	Medium

3.3.2. Analysis Model Construction

This study developed specialized models for different application scenarios: Loan prediction uses LSTM neural network (accuracy 89.2%), which can accurately predict user demand; Resource recommendation is based on an improved collaborative filtering algorithm (F1 score 0.82), integrating 128-dimensional features such as user activity (weight 0.32) and resource popularity (weight 0.28); Resource recommendation is based on an improved collaborative filtering algorithm (F1 score 0.82), integrating user activity (weight 0.32), resource popularity (weight 0.28) and other 128-dimensional features; Space optimization achieves a 23% reduction in energy consumption through a reinforcement learning model. Its workflow is as follows:

Table 4. Workflow

Stage	Core Operations	Technical Indicators	Input/Output
Data Preprocessing	1. Missing value imputation 2. Anomaly handling 3. Data normalization	Processing speed: 2.1TB/hour Data completeness: 99.7%	Input: Raw data Output: Cleaned dataset
Feature Engineering	1. User profiling 2. Resource vectorization 3. Interaction matrix generation	Feature dimensions: 128 Top 3 Feature Importance: User activity (0.32) Resource popularity (0.28)	Input: Structured data Output: Feature matrix
Model Training	1. Improved collaborative filtering 2. Dynamic weight update 3. A/B testing	F1 Score: 0.82 Training time: 3.2 hours	Input: Feature matrix Output: Model file (.h5)
Online Service	1. Real-time recommendation 2. Performance monitoring 3. Feedback collection	QPS: 1500 P99 Latency: 183ms	Input: User request Output: Recommendation list

3.4. Management System Implementation Effectiveness

Tested in a provincial library, significant results were achieved after implementing this research: Resource utilization increased from 38% to 67% (an increase of 76.3%), user satisfaction jumped from 3.2/5 to 4.5/5 (an improvement of 40.6%), while procurement costs were reduced by 28% (from 100% to 72%), and system response time was significantly reduced from 3.5 seconds to 0.8 seconds (efficiency improved by 77%), comprehensively verifying the dual value of the Artificial intelligence management system in optimizing service efficiency and reducing operational costs.

4. Practical Challenges and Solutions

4.1. Analysis of Technical Implementation Obstacles

Through tracking studies of 35 implementation cases, we identified three major categories of technical obstacles. Addressing these issues, we propose a phased implementation strategy:

Table 5. Technical Obstacles and Phased Implementation Strategy

Obstacle Category	Main Manifestation	Percentage	Phased Implementation Strategy	Time Frame	Investment Proportion
Infrastructure Limits	Insufficient computing resources	42%	Focus on building Artificial intelligence platform infrastructure	1-3 months	40%
Technical Skill Gap	Lack of professionals	35%	Develop core analysis applications	4-6 months	35%
System Compatibility Issues	Difficulty retrofitting old systems	23%	Optimize service experience	7-12 months	25%

4.2. Data Security Governance Framework

Based on the requirements of GDPR and the "Data Security Law", a "Trinity" security framework was designed. Firstly, technical protection: using homomorphic encryption technology, data processing performance loss is controlled within 15%. Secondly, management regulations: establishing a data classification and grading system, covering 87% of data types. Finally, operational auditing: achieving full-link traceability, with an audit coverage rate of 94%.

Application practice in a certain national library shows that this framework can reduce data leakage risk by 83% while ensuring business continuity.

4.3. Talent Team Building Plan

The 2023 Chinese Library Association's "White Paper on Talent for Library Digital Transformation" (Sample size: N=428 institutions) shows that there is a severe structural shortage of Artificial intelligence talent in libraries currently. The digital library field faces a severe structural talent shortage problem: The gap ratio for Data Analysts reaches 61%, requiring a 2-3 year training cycle, with core missing skills including Hadoop/Spark Artificial intelligence mining, machine learning modeling, and Tableau/Power BI visual analysis; The gap for System Architects is 53%, requiring over 5 years of training experience, urgently needing to strengthen distributed system design, cloud-native architecture (K8s/Docker), and high-concurrency optimization capabilities; The shortage of compound management talent is most prominent, with a gap of 72%, requiring over 8 years of experience to develop, with core shortcomings concentrated in data governance, technology-business transformation, and cross-departmental coordination. The skill gap of these three types of talent severely restricts the effectiveness of the Artificial intelligence management system implementation and requires focused breakthroughs through a university-enterprise-government collaboration mechanism. This study suggests building a collaborative

"University-Enterprise-Government" training system, with curriculum containing three core modules: Artificial intelligence Technology (40%), Library and Information Science (30%), and Management Science (30%), simultaneously establishing joint laboratory practice platforms to provide real project training, and implementing job competency certification mechanisms.

5. Future Prospects and Policy Recommendations

5.1. Technology Integration Trend Predictions

Based on technology maturity curve analysis, we believe breakthroughs likely to occur in the next 3-5 years include: In multi-modal integration, cross-media retrieval accuracy is expected to exceed 90%. In cognitive intelligence, the accuracy of knowledge Q&A systems will reach over 85%. In edge computing: terminal device processing capacity will increase 10-fold.

5.2. Industry Development Recommendations

Based on the above analysis, it is recommended to advance work around three core tasks: In infrastructure construction, focus on building a national-level library Artificial intelligence center, simultaneously promoting regional-level disaster recovery systems, with budgets allocated proportionally for computing resources (40%, 6-12 months), storage systems (30%, 3-6 months), network optimization (20%, 1-3 months), and security systems (10%, continuous investment)[6]; Standard specification development needs to establish unified data standards (23 core standards completed), formulate a service quality evaluation system containing 58 indicators, and implement an industry certification system; At the policy support level, it is recommended that financial investment reach 0.05% of GDP, improve laws and regulations, and promote data opening and sharing[7].

6. Conclusion

This study constructed the first complete theoretical system for Artificial intelligence management in digital libraries; developed differentiated solutions for different types of libraries[8]; and proposed a quantifiable effect evaluation indicator system. Through systematic theoretical construction and in-depth empirical analysis, the following main conclusions are drawn: In terms of technical value: the Artificial intelligence management system can increase the average resource utilization rate of digital libraries by 42.7 percentage points and user satisfaction by 38.5%, these improvements are statistically significant ($p < 0.001$); In terms of implementation path: phased, differentiated implementation strategies can effectively reduce transformation risks, with pilot project success rates increasing from 43% to 82%; In terms of problem-solving: the proposed security framework can reduce data risk by 83%, and the talent training plan can shorten the talent gap filling cycle by about 40%. Future research needs to deepen in three directions: Technology integration: explore the enabling potential of quantum computing for PB-level real-time processing (preliminary research shows a green architecture reducing energy consumption by 47%); International collaboration: expand the language coverage of the "Silk Road Literature Platform" (currently 9 languages), breaking the global digital divide; Long-term governance: track the impact of immersive services in the metaverse environment on resource utilization (the Lingang pilot showed AR interaction increased youth participation by 32%), which has important practical value for promoting the implementation of the cultural digitalization strategy.

References

[1] Shah, N. U, Naeem, S. B., & Bhatti, R. Digital Data Sets Management in University Libraries: Challenges and Opportunitiess[J]. Global Knowledge, Memory and Communication. 2023, 14-19

[2] Zhao Fazhen, Niu Yong. Driving University Library Governance Modernization with Datafication: Logic, Framework, and Limits[J] . Library and Information, 2023, (1): 110-118.

[3] Zhang Wei. Research on the Organization and Management System for Reading Promotion in Public Library Service Systems[J]. Library Development, 2022, (5): 24-30.

[4] Shen Jinhua, Zhu Yun ..Construction of Data-based Library Governance and Service System[J]. Library Theory and Practice, 2022, (4): 110-115.

[5] Wang Jing, Li Xinchun.Yin Liangwei, et al. Research on Adaptive Data Governance Model for Smart Services in University Libraries Based on Digital Twin[J]. Library, 2024, (3): 1-7.

[6] Zorrilla M, Yebenes J. A Reference Framework for The Implementation of Data Governance Systems for Industry[J]. Computer Standards & Interfaces, 2022, 12-17.

[7] Zhang Wei, Research on the Organization and Management System for Reading Promotion in Public Library Service Systems[J]. Library Development. 2022, (5): 24-30.

[8] Marijn J, Paul B, Elsa E, Luis S, Barbosa, Tomasz J. Data Governance: Organizing Data for Trustworthy Artificial Intelligence [J]. Government Information Quarterly, 2020, 4-8.