ISSN 2522-6398 Vol. 9, Issue 6: 30-34, DOI: 10.25236/FER.2026.090604

Construction and Practice of Big Data Program Laboratory in Local Universities in the AI Era

Xiaodong Tang[1,a,*]

[1]*School of Digital Economy, Hubei University of Automotive Technology, Shiyan, China*
[a]*xiaodongtang@huat.edu.cn*
**Corresponding author*

Abstract: *The rapid advancement of artificial intelligence (AI) has introduced new demands on the cultivation of big data professionals. Traditional laboratories, which often fall short in computational capacity, data resources, and scenario integration, are increasingly inadequate in the AI era. Local universities face the dual challenges of limited resources and disconnection from regional economic development. This paper proposes a laboratory construction model centered on building a general-purpose technical foundation, integrating regionally distinctive data resources, and driving instruction with real-world business scenarios. We design a three-layer cloud-terminal architecture, embed multiple industry-sector desensitized datasets, and develop a four-year progressive project-based experimental system. Taking a university-industry collaboration project at a local university as a practical case, we demonstrate the development process and application outcomes of a big data laboratory platform. Practice shows that this model effectively enhances students' AI application capabilities and their employment competitiveness in serving regional development, providing a replicable paradigm for big data laboratory construction in local universities.*

Keywords: *Laboratory Construction; Big Data Program; Local Universities; Artificial Intelligence; Industry-Education Integration*

1. Introduction

The rapid advancement of artificial intelligence (AI) is profoundly reshaping the workflow of data science. Traditional data processing and analysis workflows are gradually evolving into a new paradigm of human-machine collaboration and AI-assisted intelligence. The industry's demand for big data talent has shifted from coding and parameter tuning to a higher level of business scenario understanding, proficient use of AI tools, and complex problem-solving capabilities [1]. This transformation poses significant challenges to the cultivation system for big data professionals in higher education, particularly in the area of practical training.

Local application-oriented undergraduate institutions bear the important mission of cultivating high-quality talent to serve regional economic and social development. Taking Shiyan City, Hubei Province as an example, the city has developed a modern industrial system led by the automotive industry, with coordinated development of advantageous industries such as culture and tourism, ecological conservation, e-commerce, and logistics. However, local universities commonly face several dilemmas in laboratory construction [2]. First, computational resources are insufficient, limiting students to small-scale, low-complexity experiments. Second, data resources are oversimplified—experimental teaching mostly relies on small sample datasets accompanying textbooks, which are far removed from the industrial-grade, massive, real-time, and multi-source real data environment. Third, scenario integration is weak; experiments are mostly confirmatory and isolated, lacking comprehensive project practices that span courses, semesters, and disciplines [3]. Fourth, there is a disconnect from the local economy; experimental cases are often drawn from general domains such as e-commerce recommendation and financial risk control, showing little relevance to local pillar industries, thereby limiting students' ability to serve regional development [4].

To address the above issues, this paper aims to explore a big data program laboratory construction scheme that can both align with the frontiers of AI technology and deeply serve regionally distinctive industries. We propose a laboratory construction model for local universities, whose core lies in the organic integration of a general-purpose technical foundation, regionally distinctive data resources, and real-world business scenarios.

ISSN 2522-6398 Vol. 9, Issue 6: 30-34, DOI: 10.25236/FER.2026.090604

2. Construction Philosophy and Objectives

It is proposed that laboratory construction should adhere to integration in three aspects [5]. First, integration of technology and scenarios. Instructors should embed data collection, cleaning, analysis, modeling, and other processes into authentic industrial problems, prompting students to master technologies through project practice. Second, integration of generality and distinctiveness. The underlying technical architecture of the laboratory should maintain generality, ensuring that students master mainstream industry big data platforms and AI frameworks; the upper-level case library should incorporate local industrial data, achieving the dual goals of general capability and regional adaptability. Third, integration of human-AI collaboration. The laboratory should treat AI as infrastructure, enabling students to use tools such as AI-assisted programming and automated report generation throughout the experimental workflow, thereby cultivating human-AI collaborative data analysis capabilities [1].

Drawing on data science knowledge frameworks and local industry needs, the laboratory should support capability development from the foundational to the innovative level. The foundational level includes data collection, cleaning, storage, and SQL querying. The core level involves data analysis, mining, and machine learning modeling. The application level encompasses visual analytics and business intelligence reporting. The innovative level includes large language model (LLM) application capabilities, such as prompt engineering, retrieval-augmented generation (RAG), and agent construction [6]. The overall goal of the laboratory is to become an influential industry-education integration base for big data in the region, capable of supporting experimental teaching in multiple core courses, housing desensitized datasets from several industries, supporting students' graduation projects or innovation projects, and establishing data partnerships with multiple local enterprises [7, 8].

3. Laboratory Architecture Design

The laboratory adopts a layered architecture that balances computation, storage, networking, and management [2]. The first layer is the data source and access layer. Data sources mainly include desensitized data provided through corporate partnerships, open-source datasets, and simulated data generators. This layer can be configured with a real-time data stream simulator to generate continuous message streams, simulating scenarios such as user clickstreams or sensor data. The second layer is the platform and computing layer. This layer deploys big data storage and computing clusters using mainstream distributed storage and computing frameworks. It also builds a data warehouse for online analytical processing, configures a machine learning operations platform for model version management and training tracking, and deploys a multi-user, isolated development environment. The third layer is the AI compute layer. This layer builds a computational resource pool, configures several high-performance computing nodes, and adopts a task scheduling system for resource allocation. The laboratory can pre-deploy multiple open-source large models, supporting students' access via APIs or local invocation. An automated machine learning platform is also provided to lower technical barriers.

The core distinctive module of the laboratory is the Regional Data Supermarket. This module houses desensitized or simulated datasets from several industries, corresponding to major industries: automotive, culture and tourism, ecological conservation, e-commerce, and logistics. All data are desensitized to ensure compliant use. The automotive industry dataset contains repair work orders, fault codes, parts replacement information, etc. The culture and tourism dataset includes scenic spot ticket orders, visitor traffic, and online reviews. The ecological dataset comprises reservoir water quality monitoring data, including pH, dissolved oxygen, ammonia nitrogen, and other indicators. The e-commerce dataset consists of online sales data for local agricultural products. The logistics dataset contains waybill data from a regional logistics park. These data are obtained primarily through university-industry cooperation, government open data, and simulated generation.

The laboratory's unified portal integrates experiment management, resource requests, AI-assisted programming, and an experiment report assistant. Experiment management provides downloadable lab guides, code submission, and unit-test-based automated evaluation. Resource requests allow students to request computing resources on demand. AI-assisted programming integrates an open-source LLM interface to support natural language to code generation and error correction. The experiment report assistant allows students to input analytical conclusions and have AI generate a standardized report.

ISSN 2522-6398 Vol. 9, Issue 6: 30-34, DOI: 10.25236/FER.2026.090604

4. Laboratory Instruction and Support

It is argued that experimental projects should be structured into three levels based on cognitive progression: confirmatory, design-based, and comprehensive/innovative [3]. Confirmatory experiments aim to help students master basic tools and processes, for example, reading scenic spot visitor data and completing missing value imputation and visualization. Design-based experiments require students to solve semi-open problems and compare methods, such as comparing the performance of multiple prediction models and selecting the best. Comprehensive/innovative experiments require students to complete a full project and deliver tangible outcomes, for example, building an intelligent visitor flow early warning system that includes a front-end dashboard, a back-end prediction model, and an alerting agent [9].

To ensure the orderly conduct of experimental teaching, the laboratory needs to establish corresponding management and support mechanisms. The laboratory adopts a time-sharing strategy for compute resource scheduling [10]. During teaching hours, classroom instruction is prioritized. Outside of teaching hours, resources are open to innovation projects, competition training, and graduation projects. During idle times, research tasks are automatically run to avoid resource waste.

The laboratory adopts a project-based learning model. Starting from their first year, students choose an industry direction (one of five: automotive, culture and tourism, ecological, e-commerce, logistics) and complete a four-year progressive project around that direction, culminating in a graduation project or competition entry. This curriculum design, optimized through industry-education integration, directly targets the enhancement of students' professional competencies [11]. Taking the culture and tourism direction as an example: In the Python Programming course, students perform scenic spot ticket data cleaning and visualization. In the Data Mining course, they conduct sentiment analysis on tourist reviews. In the Machine Learning course, they build a visitor flow prediction model. In the graduation project, they develop an intelligent tour guide assistant based on LLM-powered RAG. Other directions also have clear four-year trajectories: the automotive direction moves from cleaning maintenance data to fault code mining, then to parts demand prediction, and finally to building an intelligent maintenance Q&A assistant; the ecological direction progresses from water quality visualization to anomaly detection, then to pollution diffusion simulation, and finally to a water quality warning agent; the e-commerce direction goes from order analysis to user profiling, then to sales prediction, and finally to intelligent product selection recommendation; the logistics direction moves from trajectory visualization to warehouse clustering, then to route optimization, and finally to an intelligent dispatching agent.

Regarding data security and compliance, all data in the repository have had direct identifiers removed and have been noised or generalized. Student experiments are conducted in isolated container environments where data cannot be copied externally. Data access permissions for different courses and project groups are isolated, and access to sensitive data requires instructor approval. In addition, the role of undergraduate class tutors is critical in guiding students through project-based learning in the big data era; thus, the laboratory management system incorporates tutor oversight and regular mentorship checkpoints [12]. In terms of university-industry cooperation, the university signs annual data update agreements with partner companies to ensure case library timeliness, invites corporate data engineers to conduct lab instruction or lectures each semester, and opens the laboratory to partner company employees for technical training or small projects during winter and summer breaks [8]. The university conducts end-of-semester surveys of experimental teaching satisfaction, organizes a laboratory construction seminar each semester, and updates one or two industry datasets annually, thus achieving continuous improvement [5].

In addition to traditional experiments, the laboratory has added three AI-related experimental modules. The first is prompt engineering practice, where students design prompts for maintenance work orders to extract key information from unstructured text. The second is RAG practice, where students build a knowledge Q&A system by vectorizing documents and combining them with an LLM to generate answers. The third is an introduction to model fine-tuning, where students use local computing resources to fine-tune a small open-source model for sentiment classification tasks.

5. Practical Case

A local application-oriented undergraduate university recently initiated a big data laboratory platform construction project. The project received support from a technology company through an industry-university collaborative education program, with the company providing access to a software platform

ISSN 2522-6398 Vol. 9, Issue 6: 30-34, DOI: 10.25236/FER.2026.090604

valued at several hundred thousand RMB. The two parties signed a project cooperation agreement and completed the project in phases, including hardware upgrades, software deployment, faculty training, platform pilot operation, and project acceptance.

For software platform deployment, the university selected several core software tools based on teaching needs, covering the complete data processing workflow of data collection, cleaning, analysis, mining, and visualization. At the same time, the university formulated laboratory management regulations and operating procedures to ensure standardized operation.

In terms of course application, the platform has supported the teaching of multiple undergraduate courses. Python Programming supports students in writing and debugging code online in real time. Data Mining enables students to perform association rule mining using real transaction data. Machine Learning guides students through the complete process from data preprocessing to model evaluation. Data Visualization helps students quickly generate interactive charts. Additionally, the platform has been used in a graduate-level course, *Big Data Analytics*, where instructors can demonstrate the entire data analysis process live in class, and students can practice independently after class, significantly enhancing teaching flexibility.

The platform has had a notable effect on student capability development. Using real or simulated industry data on the platform, students completed full training from data cleaning to model building, gaining exposure to enterprise-level data analysis tools and workflows while still at university. Leveraging platform resources, students actively participated in various data analysis competitions and won multiple awards at the university, provincial, and national levels.

6. Conclusion

This paper addresses the challenges faced by local universities in constructing big data laboratories in the AI era—insufficient computing power, unrealistic data, disconnected scenarios, and detachment from the local economy. It proposes a construction model that integrates a general technical foundation, regionally distinctive data resources, and real-world business scenarios, and validates the model through an industry-university collaboration project. The main conclusions are as follows. First, big data laboratories in local universities do not need to pursue comprehensive scale blindly; rather, they should integrate regional industrial characteristics into a general architecture to maximize benefits with limited resources. Second, a four-year progressive project-based learning approach helps students develop a competency structure that combines horizontal skills across the entire data workflow with vertical depth in specific industry logics. Third, using large models as infrastructure enhances students' adaptability to AI-native work patterns. Fourth, the core of industry-education integration lies in two-way interaction: laboratories provide intellectual support and talent reserves for industries, while industries provide real data and application scenarios for laboratories, supporting each other.

Based on the above findings, we suggest deepening future work in four areas: building a regional industrial big data innovation center to interface with government public data; developing an AI teaching assistant system to support experimental instruction; establishing an inter-university experimental alliance to share data and computing resources; and tracking technological frontiers by continuously introducing new technologies such as multimodal large models, agent frameworks, and AutoML to keep laboratory content evolving in sync with industry needs.

Acknowledgements

This study is supported by the Doctoral Research Initiation Fund Project of Hubei University of Automotive Technology (BK202440), the Ministry of Education University-Industry Collaborative Education Program (250600116245053).

References

[1] Rasul T, Nair S, Kalendra D, et al. The role of ChatGPT in higher education: benefits, challenges and research directions[J]. Journal of Applied Learning & Teaching, 2023, 6(1): 41-52.
[2] Han T. Research on the informatization construction of laboratories in application-oriented undergraduate colleges and universities[J]. China Management Informationization, 2024, 27(02): 224-226.

ISSN 2522-6398 Vol. 9, Issue 6: 30-34, DOI: 10.25236/FER.2026.090604

[3] Sun K, Deng X, Wang J. Research on Practical Teaching System of Data Science and Big Data Technology under the Background of New Engineering[J]. Journal of Higher Education, 2023, 9(14): 5-8.
[4] Ma J. Analysis of the Collaborative Education System of Local Universities under the Fundamentals of "New Engineering" Construction[J]. Journal of Contemporary Educational Research, 2021, 5(5): 81-85.
[5] Kui X, Kang S, Xiao Y. Research on AI-empowered innovative talent cultivation and practice in new liberal arts: A case study of the digital publishing major[J]. Industry and Information Technology Education, 2025, (07): 62-67+79.
[6] Li L, Zhang L, Liu C, et al. Teaching reform and practice of AI-empowered applied talent cultivation[J]. Economic and Social Development Research, 2025, (6): 0259-0261.
[7] Kassenkhan A M, Moldagulova A N, Serbin V V. Gamification and Artificial Intelligence in Education: A Review of Innovative Approaches to Fostering Critical Thinking[J]. IEEE Access, 2025:13.
[8] Jia L. Innovation and Practice of Teaching Reform Model of Big Data Specialty Driven by Integration of Production and Teaching[J]. Education Reform and Development, 2025, 7(2): 234-242.
[9] Yang F, Liu S, Xu H. Exploration of Building a Talent Training System for Big Data Management and Application under the Background of New Liberal Arts[J]. The Guide of Science & Education, 2024, (03): 23-25.
[10] Zhang Y. Challenges and Countermeasures of Higher Education Management in the Big Data[J]. China Modern Educational Equipment, 2023, (17): 7-9.
[11] Wang C, Zhao X, Jiang H, et al. Curriculum Optimization for the Big Data Major through Industry-Education Integration Oriented toward Enhancing College Students' Professional Competencies[J]. Curriculum and Teaching Methodology, 2024, 7(8): 30.
[12] Jiang P. Challenges and Countermeasures for University Undergraduate Class Tutors in the Big Data Era[J]. China Metallurgical Education, 2026(01): 97-99+103.