

Design and Development of an AI-Based Teaching Platform for University Physics Experiment Courses

Yuguo Luan*

School of Science, Shenyang Aerospace University, Shenyang, 110135, China

*Corresponding author: yuguo@sau.edu.cn

Abstract: To address the two major challenges prevalent in current university physics experiment instruction—namely, students' insufficient proficiency in experimental data processing and the relatively high cost of intelligent transformation—this paper designs and develops an auxiliary teaching platform for physics experiment courses based on retrieval-augmented generation (RAG) technology and a lightweight API invocation strategy. The platform adopts a hybrid technical architecture combining a local knowledge base with cloud-based API services. Centered on a self-constructed domain-specific knowledge base for physics experiments, and incorporating a lightweight invocation strategy for the Tongyi Qianwen (Qwen3.6-Plus/Max) API, the application system is deployed on existing university servers. The platform features core functionalities including intelligent Q&A on experimental principles, automated experimental data processing, intelligent assisted generation of experimental reports, and real-time diagnostic analysis of student learning progress. Furthermore, pilot teaching applications demonstrate that, with minimal additional cost, the platform effectively enhances students' experimental data processing capabilities and autonomous inquiry skills, thereby offering a replicable and affordable AI-empowered solution for experimental teaching in financially constrained local institutions.

Keywords: university physics experiment; artificial intelligence; RAG technology; lightweight API invocation; experimental teaching platform; hybrid architecture

1. Introduction

1.1 Research Background

The Ministry of Education launched a strategic action called National Education Digitalization in 2025, and the State Council's "AI+ Action Opinions requires integrating AI into every element of education and teaching, transforming talent cultivation from knowledge transmission to ability enhancement. Under such an initiative, the physics experiment teaching will shift from "traditional operation" to "intelligent inquiry". As one of the basic compulsory courses for science and engineering majors, College Physics Experiment bears the responsibility of cultivating scientific literacy, practical ability and innovative thinking. Two major pain points exist in the teaching process as follows: Students focusing on reporting rather than reflecting methods result in lack of ability transfer; Intelligent upgrade was stopped due to high cost especially for local colleges and universities who can't afford the long-term costs of high-performance GPU clusters or commercial APIs.

1.2 Current Research Status at Home and Abroad

The early research is abroad: MIT applied computer vision for tracking experimental operations; the University of Oxford has developed an AI Agent which can autonomously perform quantum experiments using a "high-end, precision, and advanced" approach. Breakthroughs domestic such as building the PhysTrack system by Peking University, developing ChatPED at Shanghai Jiao Tong University, holding a smart course launch meeting from the University of Science and Technology Beijing, all verify the potential of applying AI to improve the quality of experimental teaching. Most of these rely on expensive specialized equipment or cloud computing power leading to high cost of deployment, great privacy risk and difficult popularization.

1.3 Research Objectives and Significance

Our goal in this study is to find a "low-cost, highly available, and reproducible" reform path by building a vertical domain knowledge base on physics experiments using RAG technology and a lightweight API invocation strategy that could be deployed into existing servers. Our main idea is as follows: our local knowledge base deals with answers to high-frequency questions reducing API calls; we use an intelligent routing switch between mid-range and high-end models to control cost. It offers a replicable "democratized" reform model for universities with limited funding.

2. Overall Architecture Design of the Platform

2.1 Design Principles

Our design principles are as follows: cost controllable (average annual cost can be controlled in range from 5,000-8,000 yuan with a hybrid architecture of "local knowledge base + lightweight API invocation"), professional (dedicated knowledge base built for solving "hallucination" problem of general-purpose large models), secure (sensitive data stored locally; upload only desensitized query text into the cloud) and easy-to-use (multi-terminals access by browser or WeChat QR code scanning).

2.2 System Architecture

We adopt a hybrid approach ("knowledge enhancement + intelligent routing + tool execution") (Figure 1), which addresses the shortcoming of using one model for both complex reasoning/data processing, while having cost control^[1-3].

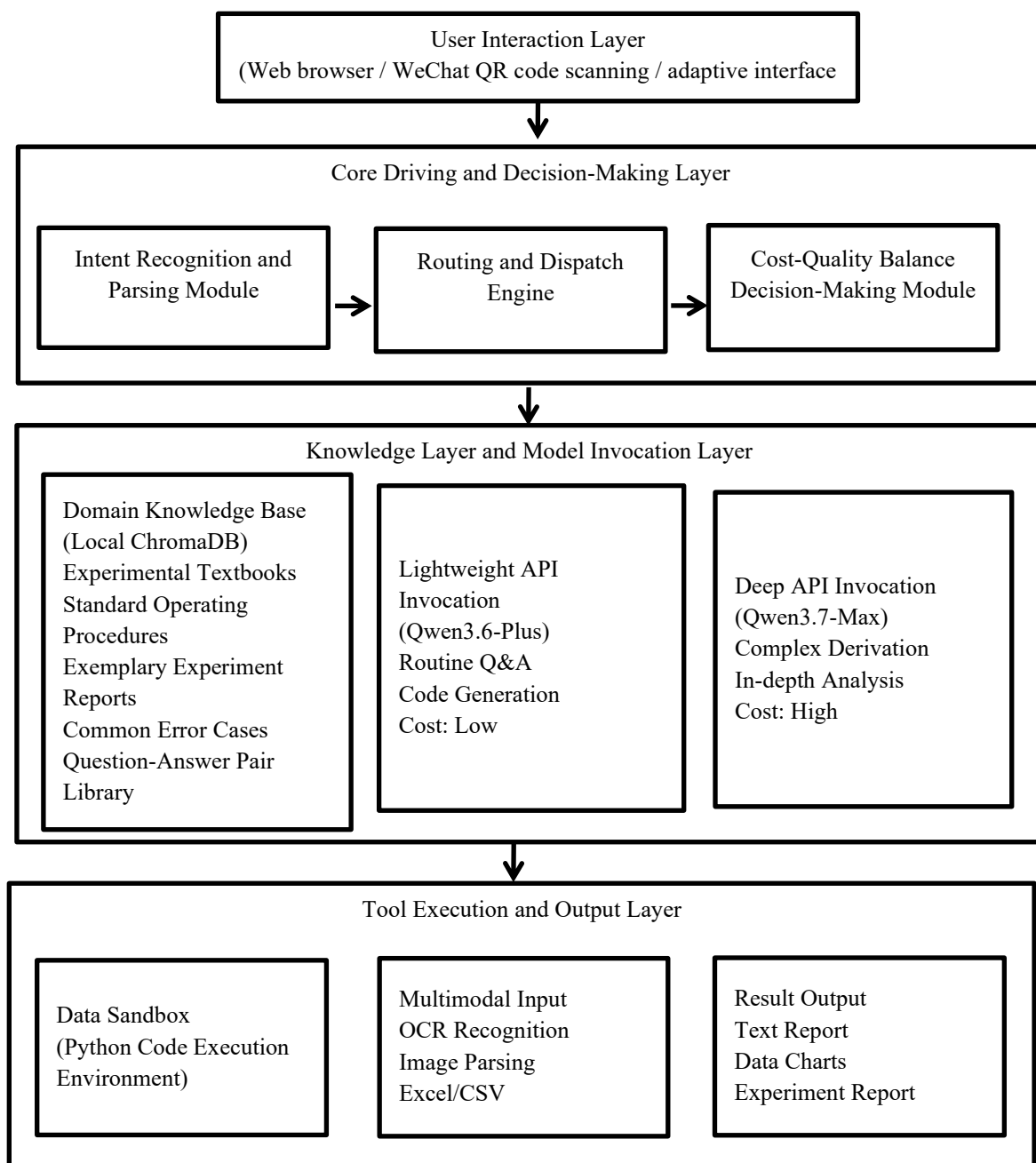


Figure 1 Overall Architecture Diagram of the Platform

2.3 Technology Selection and Implementation

We utilize Streamlit in the front end which can be adapted into WeChat Mini Programs, and LangChain as our back-end connecting with the ChromaDB vector database and cloud-based large model APIs. There are 3 levels of calling when invoking models: directly answer based on local knowledge base (zero cost), Qwen3.6-Plus handles routine Q&A (low cost) and Qwen3.7-Max used for complex derivation (high-frequency and low-usage). Embedding model is BGE-M3. Runs on existing servers where API call will be sent via HTTPS encryption.

3. Core Functional Module Design

The platform designs four major modules around the entire experimental process, forming a closed loop of "preview—operation—report—feedback."

3.1 Intelligent Parsing and Alignment of Experimental Objectives

Automate parsing the teaching syllabus, explain the objective and core skills in an experiment by using RAG to retrieve course objectives and generate introduction; List up assessment points as well as progressive relationship of knowledge between previous and next experiments, formulating a systematic cognitive framework.

3.2 Visualization of Experimental Principles and In-depth Q&A

Layered explanations on obscure principles: Basic layer: Analogies; Advanced layer: Derivation of formula/theoretical foundation; Interactive layer: Multi round of follow up questions/dynamic reasoning. Heuristic guidance: Clues/counter-questions will be provided if there's a difficulty encountered in order to guide students to think independently.

3.3 Interactive Guidance for Experimental Content (Virtual Teaching Assistant)

Acts as a real-time online "virtual teaching assistant", pushing operation suggestions according to preview and experiment progress; Inferring the reason of data deviation from the knowledge base with troubleshooting suggestions; Realizes bidirection tracing among "problem manifestations" and "knowledge blind spots", correctly locate learning obstacles^[4-6].

3.4 Automated Processing and Analysis of Experimental Data

OCR recognition of handwritten data, parsing Excel/CSV; auto identifies outliers and generates Python scripts for processing; auto computes errors & standard deviations and generate high-definition charts compliant with academic standards; and it generates a full "Experimental Result Analysis" text which converts in one click unstructured data into structured charts.

4. Knowledge Base Construction and Application of RAG Technology

4.1 Construction of a Dedicated Knowledge Base for Physics Experiments

Knowledge base: Foundation of professional AI response control cost. Build up a resource repository including different types of resources like experimental textbooks, grading standards, common errors, exemplary sample reports, academic literature, question-answers pair etc., following the strategy of semantic chunking + overlap retention(each chunk has 300 - 500 characters with overlap of 50 characters), embed them into 768 dimensional vectors by using BGE-M3 and store it locally in ChromaDB to ensure the integrity of physics formulas and derivation process.

4.2 RAG Retrieval-Augmented Generation Mechanism and Cost Control

To alleviate the "hallucination" problem of large models in specialized scenarios and control API invocation costs, the platform adopts a four-tier response strategy, as shown in Table 1.

Table 1: Four-Tier Response Strategy

Tier	Trigger Conditions	Handling Method	Cost Characteristics
Tier1	High-frequency common questions, with a Top-1 similarity ≥ 0.85	Direct Matching and Return by the Local Knowledge Base	Zero API cost
Tier2	No match is found in the Q&A library, but relevant context is retrieved	Generation by Qwen3.6-Plus Based on RAG Context	Approximately 60% - 70% of the cost of pure model generation
Tier3	Retrieval results are insufficient, but the question falls within the course scope	In-depth Analysis by Qwen3.6-Plus	Standard API cost

Tier4	None of the first three tiers is satisfied, or interdisciplinary analysis is involved	Fallback Response by Qwen3.7-Max	Accounting for approximately 5% - 8% of invocations
-------	---	----------------------------------	---

4.3 Cost-Benefit Analysis

To evaluate the economic benefits of our platform, we compared three different technical solutions (total number of students: 3,471 this semester; total token volume: 510 million) based upon:

Table 2: Cost Comparison of Different Technical Solutions

Solution	Cost Item	Unit Price (Estimated)	Annual Cost (Estimated)	Three-Year Total Cost (Estimated)
Solution A: Pure High-End API (Qwen3.7-Max)	Input Token	18yuan/million Token	5580yuan	49140yuan
	Output Token	54yuan/million Token	10800yuan	
Solution B: Pure Mid-Range API (Qwen3.6-Plus)	Input Token	2yuan/million Token	630yuan	9090yuan
	Output Token	12yuan/million Token	2400yuan	
Solution C: Local GPU Server (RTX 4090)	One-time Purchase	15000-18000yuan	—	24000-34000yuan
	Annual Electricity and Maintenance Costs	2000-3000yuan		
Solution Proposed in This Study (RAG + Tiered API)	Knowledge Base Hits (Zero Cost)	2000-3000yuan	5000-8000yuan	15000-24000yuan
	Qwen3.6-Plus	3000-5000yuan		
	Qwen3.7-Max			

It can be seen from our analysis that our proposed solution in this study costs on average 5,000 - 8,000 yuan per year (range: 15,000 - 24,000 yuan) saving about 51% as compared to the pure high-end solution; it is only slightly higher than the pure mid-range solution but has greatly improved the quality of responses particularly for complex physics derivation scenarios.

4.4 Continuous Optimization Mechanism of the Knowledge Base

We have built a closed loop of "student one-click feedback -- teacher review and confirmation -- Q&A library update" that avoids repeating mistakes on similar questions; We often analyze API invocation logs, add those frequently invoking high-end models into our local knowledge base in a virtuous cycle: affordable to use → effective to use → economical to use.

5. Teaching Application and Implementation Effects

5.1 Innovation in Teaching Models

The platform is integrated into the entire process of "preview—operation—report", forming a new teaching closed loop of "AI-assisted preview, teacher-led key explanation, AI preliminary evaluation + teacher review".^[7-10]

Table 3: Comparison of Different Teaching Models

Teaching Stage	Traditional Model	AI-Empowered Model	Role Division
Pre-class Preview	Students read printed textbooks and have limited understanding.	AI intelligent guided reading + principle visualization + preview assessment	AI-led, teacher-supervised
Classroom Operation	Teachers provide itinerant guidance, making it difficult to achieve full coverage.	AI real-time Q&A + anomaly diagnosis + step reminders	Teacher-led, AI-assisted
Data Processing	Students perform calculations manually, which is prone to errors.	AI automated processing + chart generation + error analysis	AI-led, student-confirmed
Report Writing	Formats are non-standard, and the depth of analysis is insufficient.	AI framework generation + teacher personalized review	Human-machine collaboration
Post-class Feedback	Teachers take a long time to grade (2 - 3 weeks).	AI instant preliminary evaluation + learning data visualization	AI real-time, teacher in-depth

5.2 Implementation Effect Analysis

We've deployed it on our school server and ran a full process testing (effectively completed 39 person-times), showing that: - System response accuracy/stability meets expectation; - Technical path of local knowledge base + lightweight API invocation is possible; - Average response time was <1s for direct knowledge base hit, 3-5s for RAG-augmented generation, 5-15s for API invocation.

5.3 Platform Interface Display

Main functional interfaces: homepage (list of experiment projects, learning progress tracking, personal learning dashboard, etc), experimental details page (experimental objectives, principles, operations, precautions, etc.), AI dialogue window (multi-turn dialogue, formula rendering, image upload and recognition), data processing center (data entry, automatic calculation, chart generation, etc) and teacher management backend (class learning statistics, batch grading of assignments, etc.). There are 5 major functional interfaces in our platform.



Figure 2 User Interface Display Diagram



Figure 3 Experimental Data Processing and Automatic Scoring Diagram

6. Innovation Points and Features

6.1 Innovation in the Cost Control Model of a Hybrid Architecture

We go beyond "local deployment" vs "cloud invocation", proposing a new hybrid architecture: "local knowledge base + tiered API invocation". With leveraging RAG technology, we can answer up to 35% - 40% high-frequency queries locally without any cost, and route others intelligently. Average annual cost is kept at low level (5,000 - 8,000 yuan), which provides a viable solution for institutions with limited funding^[11].

6.2 Innovation in RAG Technology for Vertical Domains

We build a specialized knowledge base for physics experiments using a four-tiered approach (local direct answer -- RAG augmentation -- lightweight API -- high-end API). All generated content is limited by our teaching syllabus and academic norms to maintain its professionalism and accuracy.

6.3 Innovation in the Human-Machine Collaborative Teaching Process

AI's contribution to standardization of knowledge transmission and initial assessment, teacher's contribution for higher order thinking development and creative guidance: Their cooperation turns knowledge transmission to ability enhancement.

7. Discussion and Limitations

7.1 Explanation of the Technical Architecture

We use a hybrid architecture of "local knowledge base + cloud API" which runs our applications and databases locally while calling Qwen API through HTTPS. We explain here: Local deployment of a 7B model needs a GPU server with costs from 10,000 - 25,000 yuan (insufficient inference speed when calling the API solution running on existing servers), sensitive information stored in local storage, upload texts are desensitization processed, no user data used by Alibaba Cloud for training; Reduced the invocation volume by about 40%, high-end models occupy 5%-8%; The annual average cost ranges from 5,000-8,000 yuan^[12].

7.2 Current Problems

API dependency risk: Policy change of network/service providers will disrupt services. Use a primary (Alibaba Cloud) and Zhipu AI as backup but we need to improve our switching mechanisms; Insufficient local computing power: No independent GPU, response time extends for 15-20 seconds in peak periods Sample limitations: Only two mechanics experiments covered by 80 students so that there are insufficient statistics, further verification should be done.

7.3 Optimization Strategies

For the above problems, our follow-up strategies are: improve automatic switching among multiple service providers and local caching, explore CPU inference of Qwen2.5-7B as emergency backup; Apply for GPU resources, introduce asynchronous processing and request queues, convert non-real time tasks into background batch processing; Expand to electromagnetism and optics experiments, introduce standardized scales and third-party evaluation, do an A/B test on comparing the difference in performance between a single model solution vs the RAG solution.

8. Conclusion

Inspired from the idea of "low-cost, highly available, and reproducible", we propose a hybrid architecture of "local knowledge base + tiered API invocation" A new AI-empowered physics experiment teaching platform that at an extremely low average annual cost (5,000 - 8,000 yuan) alleviates both pain points: weak data processing capability on one hand; high investment in intelligent upgrade on the other hand, provides a replicable reform model for universities with limited funding.

Fund Projects

This paper is a phased achievement of the 2026 university-level teaching reform project "Research and Practice on the Construction of an Intelligent Course for AI-Empowered College Physics Experiments" (JG261404C2).

References

- [1] Ministry of Education. 2025 Deployment Meeting of the National Education Digitalization Strategic Action. 2025.
- [2] State Council. Opinions on Deeply Implementing the "AI+" Action. 2025.
- [3] Chen Hongzhang. Innovation in AI-Empowered Information-Based Teaching Models in Higher Education and Optimization of the Evaluation System. *Journal of Hubei Adult Education Institute*, 2026, 32(3): 14-20.
- [4] Meng Yakun, Sun Jinghao, Xue Fanghong, et al. Path and Practical Exploration of AI-Empowered Smart Course Construction. *Laboratory Science*, 2025, 28(5): 68-70.
- [5] He Chunying, Wang Yu, Guo Jing. Construction and Practice of an AI-Empowered Smart Teaching Model: A Practical Study on College English Course Teaching Reform Based on a Digital-Intelligent Teaching Platform. *Technology Enhanced Foreign Language Education*, 2025(2): 62-64.
- [6] Shen Huoming, Gong Hui, Fu Haiying. Construction of Smart Courses: Innovative Transformation Path from Digital Empowerment to Teaching Models. *China University Teaching*, 2025(9): 10-17.
- [7] Liang Zhumei, Li Bao, Zhao Dongmei. Reconstructing the Learning Process with AI Agents: Case Analysis and Reflections on Creating Teaching Agents. *China University Teaching*, 2025(9): 80-86.
- [8] Li Junfeng, Xiao Gensheng, Wei Chun'ao, et al. Construction and Practice of a "Color Science" Smart Course Based on a Knowledge Graph. *Printing and Digital Media Technology Study*, 2025(3): 142-148.
- [9] Wu Zihao, Li Ning. Exploration and Practice of Teaching Reform in the "Theory of Machines and Mechanisms" Course Based on a Knowledge Graph. *South Agricultural Machinery*, 2026, 57(S1): 168-171.
- [10] Xie Youru, Lu Yi, Peng Zhiyang, et al. Exploration of Knowledge Graph-Empowered Integration of "Teaching-Learning-Assessment" in University Courses. *China Educational Technology*, 2024(12): 1-7.
- [11] Zhou Kun, Zhu Yutao, Chen Zhipeng, et al. YuLan-Chat: A Large Language Model Based on Multi-Stage Curriculum Learning. *Chinese Journal of Computers*, 2024.
- [12] Pei Rong. Generative Artificial Intelligence Empowering Education and Teaching: Transformative Impacts, Risk Challenges, and Practical Paths. *Contemporary Education Forum*, 2024.