# Exploration and Practice of Reforming Experimental Teaching in Universities Empowered by Large Language Models

## -- Case Study of the Integrated Virtual Simulation Training Course on Modern Enterprise Operations

Qin Zhang[1], Zhongqin Xie[2, *], Huimei Wang[1]

[1] School of Management, Wuhan College, Wuhan, Hubei 430212, China

[2] Office of Academic Affairs, Wuhan College, Wuhan, Hubei 430212, China

*Corresponding author

## Abstract

**Based on the concept of OBE (Outcome-Based Education), this paper innovatively constructs an experimental teaching system for management students. Taking the integrated virtual simulation training course on modern enterprise operations as an example, it develops and implements a "four-stage progressive" experimental teaching system empowered by large language models from the perspective of teacher-student-AI triadic collaboration. Through full-chain reconstruction across "cognitive empowerment-design empowerment-implementation empowerment-evaluation empowerment," large language models are deeply integrated into every stage of experimental teaching, forming a new teaching ecology characterized by intelligent guidance, dynamic generation, human-AI collaboration, and precise assessment. The model has been practiced in training courses offered across six undergraduate majors, including big data management and application, supply chain management, business administration, and logistics management. Teaching practice shows that the introduction of large language models has enabled a shift from knowledge transmission to intelligently enhanced competence cultivation, thereby improving students' ability to adapt to the intelligent era.**

## Keywords

**Large language models, teacher-student-AI triadic collaboration, OBE, AI agents, experimental teaching.**

## 1. Introduction

Since 2022, large language models (LLMs), represented by ChatGPT, have triggered a new wave of artificial intelligence revolution. Their powerful capabilities in natural language understanding, logical reasoning, and content generation are reshaping education [1]. The Ministry of Education's 2024 Action Plan for the Digitalization of Education explicitly called for exploring innovative applications of artificial intelligence technologies in experimental teaching and promoting a transformation toward personalized and intelligent talent cultivation models [2]. LLMs represented by ChatGPT, DeepSeek, and Kimi possess strong natural language understanding and generation capabilities. Their applications cover key links such as intelligent question answering, content generation, and learning analytics, providing technical support for reconstructing the paths of knowledge acquisition and problem solving and for shifting the focus of teaching [3].

As a key vehicle for cultivating students' practical and innovative abilities, experimental teaching still faces many challenges as it explores the deep integration of emerging digital technologies with educational practice in response to the structural demand for interdisciplinary innovative talent generated by industrial transformation [4-6]. First, the contradiction between limited experimental resources and diverse student needs is prominent, and traditional fixed-menu experiments cannot meet personalized learning needs. Second, teacher guidance coverage is insufficient; in large experimental classes, students cannot obtain immediate feedback on their problems, which affects the continuity of learning. Third, experimental evaluation systems overemphasize results and lack effective means to assess higher-order abilities such as procedural innovation and critical thinking. The emergence of LLMs provides technical possibilities for addressing these difficulties. However, existing studies mostly remain at the level of theoretical discussion and simple tool use, lacking systematic model design and deep teaching-integration practice [7]. Meanwhile, the accelerated evolution of AI technologies and the upgrading of industrial capacity have placed higher-order demands on talent structures, and the concept of Outcome-Based Education (OBE) has become a core paradigm of teaching reform [8]. How to construct an LLM-enabled model suited to the laws of experimental teaching, how to handle role boundaries in human-AI collaboration, and how to evaluate its actual teaching effectiveness have become urgent practical issues. Focusing on the deep integration of LLMs and the OBE concept, this paper reconstructs the full chain of experimental teaching through cognitive empowerment, design empowerment, implementation empowerment, and evaluation empowerment, deeply embedding LLMs into every link of experimental teaching and thereby forming a new teaching ecology characterized by intelligent guidance, dynamic generation, human-AI collaboration, and precise assessment. It then conducts an empirical study through the integrated virtual simulation training course on modern enterprise operations and, in light of the latest progress in applying LLMs in professional fields, builds a practical teaching system through layered tasks.

# 2. Theoretical Framework for LLM-Empowered Reform of Experimental Teaching

## 2.1. Construction of the "Four-Stage Progressive" Empowerment Model

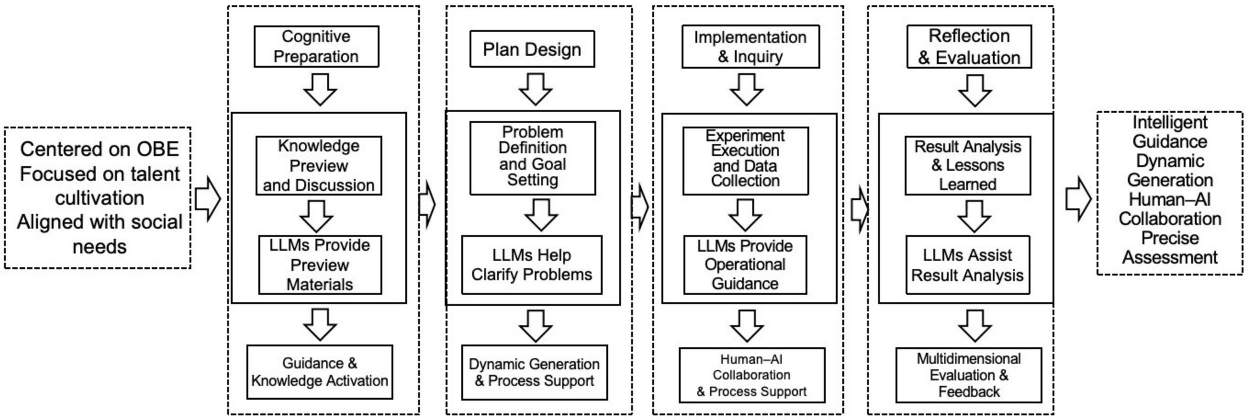


**Figure 1.** Framework of the "Four-Stage Progressive" LLM-Enabled Experimental Teaching Model

This study is grounded in constructivist learning theory and activity theory. The design philosophy of the LLM-based practical teaching framework is "centered on the OBE concept, aimed at talent cultivation, and aligned with social needs." Based on this, the study constructs a "four-stage progressive" LLM-enabled experimental teaching model (see Figure 1). This model decomposes the practical teaching process into four stages-cognitive preparation, plan design, implementation and inquiry, and reflection and evaluation-corresponding to four empowerment mechanisms.

(1) Cognitive empowerment stage. Intelligent guidance and knowledge activation. LLMs play the role of "cognitive scaffolding." Through multi-turn dialogue, they diagnose students' prior knowledge and generate personalized preview materials. For example, when a student asks, "How should I understand the difference between line changeover and investment?", the LLM not only explains the concepts, but also pushes dynamic simulation cases, guides the student to analyze key variables, and generates targeted preview questions. The core technology at this stage is the construction of a domain knowledge base based on retrieval-augmented generation (RAG), which ensures content accuracy and timeliness.

(2) Design empowerment stage. Dynamic generation and plan optimization. In traditional experiments, students often follow rigid, fixed procedures. LLMs can dynamically generate differentiated experimental tasks according to students' interests and ability levels. For example, in a logistics distribution routing experiment, based on the industry scenario selected by the student (pharmaceuticals/fresh produce), the LLM automatically generates virtual orders with different constraint conditions (timeliness, cost, temperature control) and recommends experimental paths. The design empowerment stage emphasizes "controlled generation" under teacher leadership, embedding teaching objectives and constraint rules through prompt engineering.

(3) Implementation empowerment stage. Human-AI collaboration and process support. During the experiment, LLMs act as "intelligent collaborators" and provide three types of support: (i) real-time question answering to resolve operational confusion; (ii) anomaly diagnosis to help analyze reasons when experimental data deviate from expectations; and (iii) creativity stimulation, guiding students to explore boundary conditions through heuristic prompts such as "What would happen if...?" The implementation empowerment stage adopts a Human-in-the-Loop mechanism in which key decision points must be confirmed by teachers to ensure safety.

(4) Evaluation empowerment stage. Multidimensional evaluation and enhanced feedback. LLMs analyze data generated throughout the experimental process, including operation logs, decision trajectories, and dialogue records, and generate evaluation reports across four dimensions: knowledge mastery, skill proficiency, innovative thinking, and collaborative contribution. Unlike traditional evaluation that focuses only on results, LLMs can identify students' flashes of critical thinking and innovative attempts during experiments. Even when the experiment fails, they can still provide positive evaluation, thereby protecting students' enthusiasm for exploration.

## 2.2. Teacher-Student-AI Triadic Collaborative Mechanism

The effective operation of the "four-stage progressive" LLM-enabled experimental teaching model depends on establishing a clear teacher-student-AI triadic collaborative mechanism. The role of experimental instructors shifts from "lecturer" to "designer + facilitator + evaluator," responsible for setting teaching goals, reviewing generated content, guiding higher-order discussion, and cultivating ethical awareness. Students shift from "passive executors" to "active constructors"; they need to improve AI literacy and learn how to ask questions, discriminate information, and think critically. AI is positioned as an "intelligent learning companion," with strengths in breadth of knowledge, generation speed, and tirelessness, but it must be clearly recognized that it does not possess emotional support, value judgment, or ultimate responsibility.

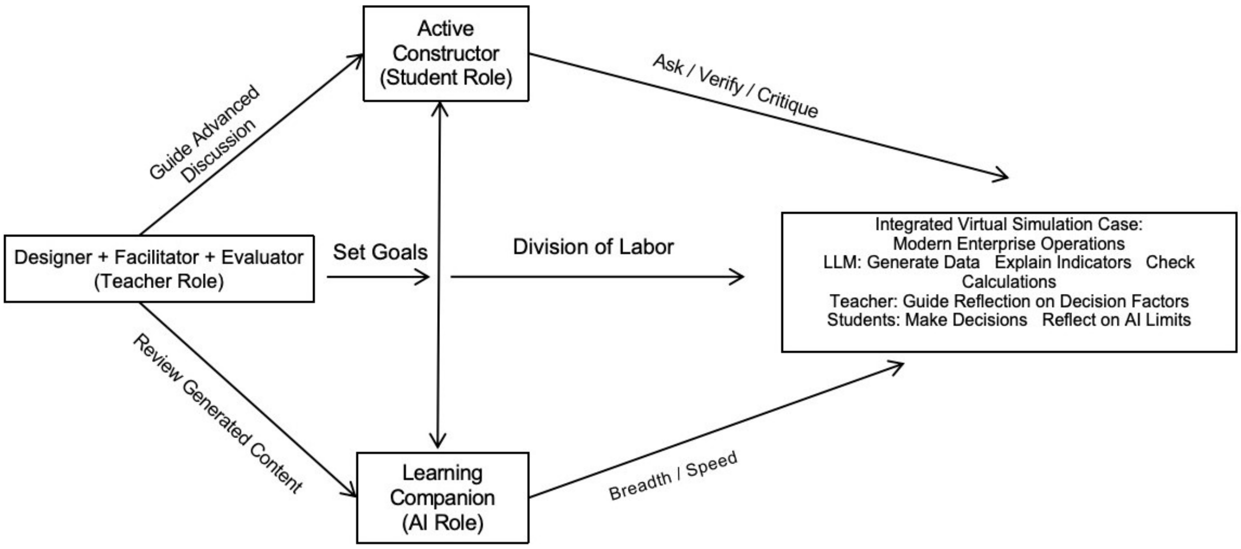


**Figure 2.** Teacher-Student-AI Triadic Collaborative Mechanism

The interaction among the three follows a division-of-labor logic in which "AI handles high-frequency repetitive tasks, teachers focus on higher-order educational activities, and students engage in deep exploratory learning." For example, in a modern enterprise operation simulation decision-making experiment, the LLM is responsible for generating decision data for the optimal human-resource allocation of production lines, explaining the meaning of indicators, and checking calculation accuracy; the teacher guides students to think about the impact of non-financial factors on decision making; and students make final decisions through comprehensive judgment and reflect on the limitations of AI suggestions.

## 3. Design of LLM-Empowered Experimental Teaching

Using the integrated virtual simulation training course on modern enterprise operations as the carrier, this paper, based on the technical characteristics of LLMs and the OBE concept, mainly targets course experiments in six undergraduate majors, including big data management and application, supply chain management, business administration, and logistics management. With support from cluster servers and student computers, open-source LLMs are deployed and interactive development environments are configured to assist students in mastering enterprise operation rules and designing comprehensive experimental decision-making skills.

### 3.1. Mastering Rules and Learning Tools

Teachers guide students on an information technology platform to complete task-based learning of theoretical knowledge points and corresponding chapter discussions and tests. This ensures that students acquire the theoretical foundation necessary for training. At the same time, LLMs are introduced to help students understand the cross-application of rules across different types of enterprises.

### 3.2. Skill Formation and Tool Reuse

Teachers guide students to master LLM questioning techniques, narrow the scope of prompts, and use LLM tools in a targeted manner for specific student needs. Students use LLMs to generate decomposed operational strategies for internal business decisions in different types of enterprises and annotate them in detail, thereby understanding the cross-application relationship between decision-making logic and concrete decision implementation.

## 3.3. Decision Optimization and Plan Evaluation

After completing the relevant training on decision-making skills, students use LLMs to parse training logs, generate visualized reports, and identify potential problems. If the decisions recommended by the LLM do not match real conditions, students revise their prompting strategy, and the LLM recommends revision plans so that errors can be corrected in a timely manner.

Through the teaching pathway of mastering rules and learning tools, skill formation and tool reuse, and decision optimization and plan evaluation, large language models are deeply integrated into experimental teaching, which not only improves teaching efficiency but also cultivates students' data thinking and innovation ability.

# 4. Project-Driven Teaching Model and Case Application

To achieve the systematic cultivation of student abilities under the OBE concept, taking the integrated virtual simulation training course on modern enterprise operations as an example, teachers adopt a project-driven teaching model featuring "problem-driven, task-oriented" instruction. LLMs are embedded into the scenarios of four stages-cognitive preparation, plan design, implementation and inquiry, and reflection and evaluation-to design a layered and progressive task system that promotes students' development from cognitive understanding to skill advancement and ultimately to quality enhancement.

## 4.1. Cognitive Preparation Stage

This stage corresponds to the "cognitive empowerment" mechanism and aims to use LLMs to build a dynamic knowledge graph, thereby addressing the pain points of complex rules and difficult interdisciplinary knowledge integration in traditional training for the integrated virtual simulation course on modern enterprise operations. One week before the training begins, students complete preparatory learning through the Kimi intelligent learning companion. For example, regarding production-line staffing, students input their knowledge blind spots, such as not understanding the relationship among "production line capacity, professional competence, and production-line staff utilization" for different types of production lines. Based on RAG technology, the LLM retrieves the course knowledge base and industry case base and generates a three-layer progressive learning package: (1) concept explanation layer: taking cold-chain transportation in the pharmaceutical industry as an example, it dynamically demonstrates the quantitative impact of safety stock on stockout risk and service level; (2) simulation warm-up layer: students are allowed to adjust Excel parameters and observe fluctuating data through optimization solving; and (3) diagnostic test layer: the system generates three situational multiple-choice questions and one open-ended reflection question. After students answer, the LLM automatically analyzes their cognitive misconceptions and feeds the results back to the teacher terminal. Teachers then adjust the focus of offline ice-breaking seminars in a targeted manner based on the class knowledge-weakness map.

At this stage, under the teacher-student-AI triadic collaborative mechanism, teachers focus on distilling common problems and providing value guidance during learning-for example, emphasizing the effective utilization of workers' capabilities. Students shift from passively watching videos on the informatization platform to actively constructing knowledge through dialogue, and their AI literacy and questioning ability begin to emerge. AI, meanwhile, completes high-frequency diagnosis, decision interpretation, and personalized resource recommendation.

## 4.2. Plan Design Stage

This stage corresponds to the "design empowerment" mechanism. Under OBE target constraints set by teachers, LLMs dynamically generate multi-gradient project tasks, realizing "different designs for different learners." Upon entering the strategic decision-making module, student teams choose an industry (for example, high-end new-energy vehicle manufacturing). Through prompt engineering, the teacher inputs the following constraints into DeepSeek: "(1) the task must cover three dimensions: capacity planning, raw-material procurement plan selection, and related decision choices; (2) at least two market-uncertainty events must be embedded (finished-goods procurement quantity/cash-flow shortage); and (3) the difficulty coefficient must be divided into three levels to ensure a match with each group's zone of proximal development." Based on these constraints, LLMs generate differentiated initial scenarios. For example, beginner groups receive structured data templates and decision checklists, with tasks focusing on basic indicator calculation; intermediate groups face semi-structured information (such as ambiguous market-demand forecasts) and need to define assumptions independently; advanced groups receive real de-identified industry data and are required to introduce ESG ratings as non-financial constraints.

At this stage, under the teacher-student-AI triadic collaborative mechanism, teachers act as "architects," reviewing the alignment between generated tasks and OBE objectives as well as ethical boundaries, to avoid generating discriminatory market assumptions. Students shift from "ordering from a menu" to "co-design," and can further question the LLM to optimize task logic, thereby cultivating needs-definition ability. AI undertakes the heavy work of combinational innovation and can generate 12 groups of heterogeneous tasks within five minutes.

## 4.3. Implementation and Inquiry Stage

This stage corresponds to the "implementation empowerment" mechanism. LLMs are embedded into the virtual simulation platform and, as "invisible teammates," provide real-time procedural support while following the Human-in-the-Loop principle to ensure the safety of key decisions. For example, in the third quarter of simulated operations, a manufacturing enterprise discovers that net profit has plunged by 18% and falls into a decision deadlock. Students initiate multi-round dialogue with the LLM: (1) anomaly diagnosis. After students upload a screenshot of the operation log, Kimi automatically identifies that "the direct cause is a surge in depreciation and amortization caused by converting to a flexible production line, while the root cause is product-positioning bias caused by insufficient earlier market research," and generates a Sankey diagram to visualize cost flows. (2) Scenario simulation. Students in a trading enterprise ask, "If we now add RMB 200,000 in advertising expenditure to improve brand premium, when will the break-even point be reached?" Based on the built-in financial model, the LLM conducts Monte Carlo simulation and returns a probability distribution rather than a single answer, warning that attention should be paid to the risk that advertising expenditure may trigger a cash-flow break. (3) Creativity stimulation. When a team proposes the aggressive plan of "dismissing management personnel," the LLM triggers a teacher alert and instead guides students by noting that layoffs may reduce costs in the short term, but may also cause severance costs to rise. The teacher should then intervene immediately in the group's decision making and guide students to choose the optimal plan.

At this stage, under the teacher-student-AI triadic collaborative mechanism, teachers monitor value deviations in AI suggestions and guide students to deepen post-failure reflection after re-decision. At nodes involving employee rights, compliance risks, and similar issues, teachers can "take over with one click." Students focus on higher-order judgment and learn to ask AI for analysis rather than answers. AI, meanwhile, handles high-frequency real-time calculation, data visualization, and pattern recognition.

## 4.4. Reflection and Evaluation Stage

This stage corresponds to the "evaluation empowerment" mechanism. LLMs collect digital footprints throughout the entire process and generate four-dimensional ability profiles that go beyond outcomes, promoting an evaluation shift from "GPA orientation" to "growth orientation."

After the training ends, experimental scores are ranked according to owners' equity, but students' reflection reports may receive higher-order evaluation from LLMs. The system analyzes human-AI dialogue records and performs decision backtracking to generate dynamic radar charts. The results show the team's knowledge mastery, skill proficiency, innovative thinking, and collaborative contribution, and produce personalized development suggestions.

At this stage, under the teacher-student-AI triadic collaborative mechanism, teachers conduct value-added evaluation based on AI reports, identify "low-score/high-potential" students, and provide recognition. Students identify ability shortcomings through visualized feedback and formulate precise improvement plans. AI, meanwhile, realizes full-sample and fine-grained process analysis.

## 5. Conclusion and Prospects

The "four-stage progressive" experimental teaching system empowered by large language models constructed in this study, together with practical verification under the teacher-student-AI triadic collaborative mechanism, shows that the model can effectively address problems such as constraints on experimental teaching resources, insufficient personalized guidance, and weak cultivation of innovation ability. It achieves dual improvement in teaching efficiency and quality and forms a full-chain empowerment mechanism of "controlled generation-process support-precise evaluation," providing a replicable practical path for experimental teaching reform in the era of generative artificial intelligence. The study still has limitations: (1) the sample is limited to management experiments, and the applicability of the model to high-risk science and engineering experiments requires further exploration; (2) long-term effects remain to be observed, and the deeper impact of LLMs on students' metacognitive ability is still unclear; and (3) ethical regulation is at an early stage and needs to be continuously improved in practice. Future research may focus on three directions: first, expanding to interdisciplinary experimental scenarios and building a domain-adaptive LLM fine-tuning framework; second, further investigating the mechanisms through which LLMs influence higher-order thinking ability; and third, exploring the establishment of co-construction and sharing mechanisms for university experimental-teaching large models to reduce the threshold for application.

## Acknowledgments

Funding: Wuhan College Teaching Reform Project (2020), "Exploration of a Blended Experimental Teaching Model Based on a Mobile Learning Platform" (WYJY202011); Wuhan College Research Project (2020), "Design of a Practice-Oriented Teaching Model for Economics and Management Courses Based on the OBE Concept" (KYY202005).

## References

[1] Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224.

[2] Ministry of Education and Eight Other Departments. (2025, April 11). Opinions on accelerating the advancement of education digitalization [J/OL]. https://www.gov.cn/zhengce/zhengceku/202504/content_7019045.htm

[3] Luo, F., Ma, Y., & Jiao, L. Z. (2025). Iteration-driven transformation: How DeepSeek's technical characteristics empower digital transformation in education. Journal of Soochow University (Educational Science Edition), 13(3), 28–39.

[4] Zhang, L. N., Gu, R. T., & Wang, Q. Y. (2025). A study on the application of large language models in language experiment teaching. Experimental Technology and Management, 42(10), 201–209.

[5] Xu, J. N., Huang, N., & Song, H. (2025). Large language model-driven automated grading of multimodal experiment reports. Modern Information Technology, 9(12), 79–84.

[6] Shao, X. X., Yang, J. X., Zhao, L. B., et al. (2025). Empowering the cultivation of elite innovative talents in fundamental disciplines at local applied universities through large language models: Challenges, opportunities, and pathways. Journal of Education and Media Studies, (2), 37–42.

[7] Liao, N. Y., Tian, Y., Li, Y. S., et al. (2025). Tool learning with large language models: Methods, functions, and mechanisms. Computer Engineering, 51(12), 1–17.

[8] Huang, L. J., Li, M. Y., & Li, X. J. (2025). Reform and practice exploration of teaching models for e-commerce specialty courses under the "OBE philosophy + AI technology" approach: Taking the "Business website construction and management" course as an example. New Curriculum Research, (35), 13–15.