ISSN: 2424-8460(O)
2251-2608(P)

Original Research Article

Research on the reconstruction and practice path of big data course teaching mode driven by generative artificial intelligence technology

Hao Sun
Communication University of China, Nanjing, Nanjing, Jiangsu, 211199, China

Abstract: This paper explores the pedagogical transformation of big data courses through generative AI, first examining traditional teaching challenges: content-industry mismatch, inadequate alignment between practice and real-world scenarios, and limitations in instructional implementation and evaluation systems. It then proposes a restructuring framework: clarifying the division of responsibilities between human and machine collaboration, developing a dynamically evolving curriculum system based on knowledge progression, and implementing precision teaching with multidimensional assessment through student analytics. Finally, case studies from Beijing No.11 School and Khan Academy demonstrate generative AI applications in data tracking, knowledge delivery, and personalized instruction, offering actionable insights for advancing big data course reform.
Keywords: generative AI; big data course; teaching model reconstruction; practical path; human-machine collaboration

1. Introduction

As global education digitalization advances, generative AI is reshaping the educational ecosystem. Yet big data courses face challenges including content-industry disconnect, insufficient practical application scenarios, and monotonous assessment methods. Exploring integration pathways between generative AI and big data courses has become crucial to address traditional teaching limitations, drive pedagogical innovation, and ultimately meet industry demands while elevating educational quality.

2. Core challenges in pre-generation AI empowered big data course teaching

2.1. Supply-demand mismatch between teaching content and industry demand

Before generative AI (GAI) integration, big data course modules remained rigidly anchored in traditional technologies like Hadoop distributed file systems and Spark computing frameworks, failing to adapt to industry-wide technological shifts. As enterprises transitioned from static batch processing to dynamic intelligent stream processing, real-time data pipeline deployment and automated model inference integration became critical business needs. However, the curriculum's knowledge refresh cycle lagged behind technological advancements[1]. The courses only covered basic data cleaning and batch processing task development, omitting essential industry technologies such as Flink stream processing frameworks, real-time data monitoring, and intelligent anomaly detection. This knowledge gap ultimately created a disconnect between students' skill sets and actual job competency requirements.

2.2. Insufficient adaptation of practical teaching to real situations

Practical training often relies on idealized cases and clean data, lacking real-world training in noise processing, multi-source integration, and resource scheduling. Tasks are typically limited to single-point technical verification, such as Hive queries and SparkRDD programming, without covering the end-to-end data engineering process in enterprises, including requirement analysis, technology selection, deployment, and operations. Students follow step-by-step procedures, making it difficult for them to encounter real-world challenges like data skew or high concurrency, resulting in insufficient ability to transfer and solve complex scenarios.

ISSN: 2424-8460(O)
2251-2608(P)

2.3. The double limitations of teaching implementation and evaluation system

Teachers are overwhelmed with explaining foundational knowledge, answering questions, and grading assignments, leaving little room to cultivate students 'data thinking and system modeling skills. Moreover, the lack of effective learning analytics tools makes it difficult to identify individual weaknesses, hindering personalized teaching. The evaluation system primarily focuses on code execution, overlooking critical dimensions like logical coherence, scalability, maintainability, and business understanding. This approach fails to align with industry competency standards and inadequately assesses students' overall capabilities.

3. Reconstruction path of big data course teaching mode driven by generative AI

3.1. Reconstructing the teaching subject and clarifying the division of rights and responsibilities between human and machine

Establish a collaborative teaching framework featuring "human-teacher-machine-assist" synergy, clearly defining the roles of educators and generative AI. Teachers are tasked with setting big data course objectives, focusing on cultivating data thinking, system modeling, and innovation capabilities. They design comprehensive tasks covering data collection, cleaning, storage, analysis, and visualization, while ensuring compliance with data privacy and technical ethics. Generative AI serves as the knowledge provider, explaining frameworks like Hadoop, Spark, and Flink through natural language interaction, breaking down algorithmic logic, and offering code generation and optimization services with real-time error diagnosis. Meanwhile, GAI handles basic data processing tasks such as unstructured data cleaning, format conversion, and repetitive statistical calculations. The system achieves intelligent regulation through a mechanism of "perceiving learning states—Inferring knowledge needs—Generating teaching services—Feedback-driven optimization," assisting teachers in enhancing instructional efficiency and precision.

Table 1. Restructuring of teaching entities and clarifying the division of responsibilities between humans and machines.

teaching subject	core rights and duties
teacher	Establish teaching objectives, cultivate higher-order competencies, design end-to-end tasks, and uphold value orientations such as data privacy and technical ethics.
production AI	Analyze technical principles, answer programming questions, provide code services, complete basic data processing, and establish a closed-loop teaching system to support instruction.

3.2. Reconstructing the course content to build a dynamic and evolving knowledge system

The generative AI-powered system builds a dynamically updated big data course knowledge base, structured into four core modules: Data Science Fundamentals, Machine Learning Algorithms, Big Data Engineering Practices, and Intelligent Project Development. In the Data Science Fundamentals module, the Generative AI (GAI) derives statistical formulas, clarifies concepts, and automatically updates cutting-edge research. The Machine Learning Algorithms module utilizes GAI to generate diverse training scenarios, deconstructs algorithmic logic and implementation steps, and provides performance optimization strategies. For the Engineering Practices module, GAI tracks industry technological advancements, continuously integrating practical content such as new big data framework features, real-time processing, resource scheduling, and troubleshooting. The Intelligent Project Development module employs GAI to break down requirements and select technologies, aligning with real-world project workflows and standards to ensure continuous synchronization between course content and industry demands.

3.3. Reconstruction of teaching process and evaluation system to implement precision teaching and multi-dimensional evaluation

In terms of teaching processes, a tiered instruction mechanism is established by leveraging generative AI's learning analytics capabilities. The GAI collects students' code records, knowledge point responses, and project implementation data to identify learning gaps, then automatically delivers tiered resources including foundational reinforcement, skill enhancement, and innovative expansion. The practical component adopts a "decreasing support" strategy: initially providing basic frameworks, mid-term offering guided thinking prompts, and later focusing solely on problem types to gradually cultivate students 'self-solving abilities. Meanwhile, the GAI

ISSN: 2424-8460(O)
2251-2608(P)

handles administrative tasks such as scheduling, attendance tracking, and preliminary assignment evaluation to improve teaching efficiency[3]. Regarding assessment systems, a multi-dimensional evaluation model is constructed covering knowledge mastery, code logic, system scalability, maintainability, modeling innovation, business understanding, and collaboration skills[3]. The GAI generates objective evaluation reports based on learning process data, while teachers provide subjective assessments incorporating students' performance in innovation and complex problem-solving, forming a collaborative evaluation mechanism of "GAI objective assessment + teacher subjective evaluation" (see **Figure 1**).

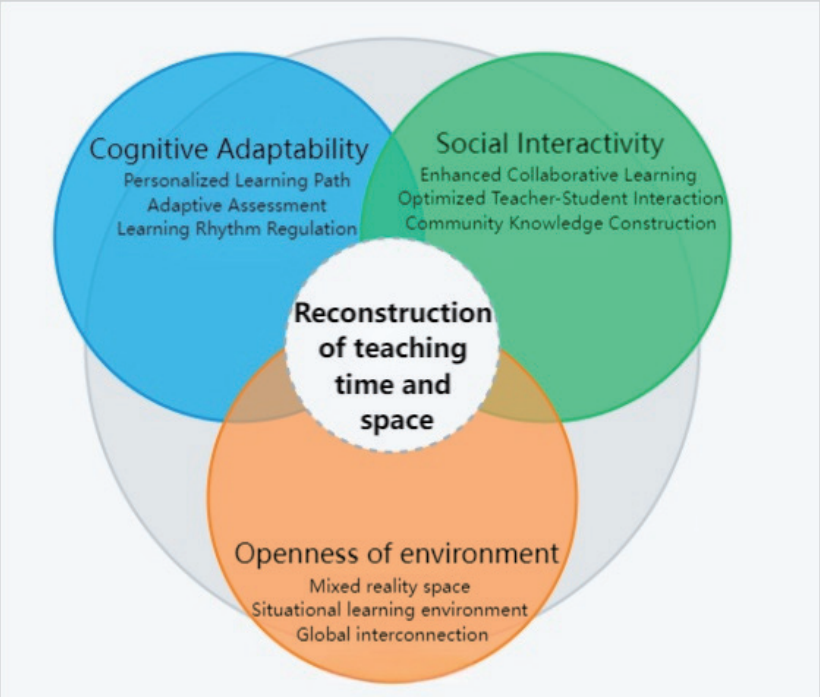

Figure 1. AI-powered multidimensional assessment of teaching competencies.

4. Case study of generative AI-driven big data course teaching

4.1. Practice of big data course intelligent teaching system in Beijing No.11 School

Beijing No.11 School utilizes a generative AI teaching system in its big data curriculum, dynamically collecting and analyzing student data from code submissions, knowledge point responses, and practical project operations. The system accurately identifies students' weaknesses in Hadoop deployment and Spark programming, providing timely feedback to instructors. During instruction, teachers implement a comprehensive workflow for data collection, cleaning, storage, and analysis, while reinforcing education on data privacy and technical ethics[4]. The generative AI handles repetitive tasks such as explaining framework principles, analyzing algorithm logic, automatically generating and optimizing code, real-time error diagnosis, and processing unstructured data. It also automatically recommends tailored resources based on individual learning progress, facilitating personalized teaching.

4.2. Teaching practice of generative AI in big data course of khan academy

Khan Academy has developed an online intelligent teaching system powered by generative AI, which structures abstract concepts like distributed computing and data modeling to simplify learning. The system monitors students' progress in real time and provides step-by-step guidance through generative AI when issues like data cleaning or machine learning algorithm applications arise, helping learners develop clear problem-solving approaches. It also generates personalized learning paths based on data, delivers advanced resources, and supports students in building their own big data knowledge systems. Additionally, the system produces learning analytics reports that help teachers refine teaching strategies and resource allocation, thereby enhancing instructional precision and learning outcomes.

ISSN: 2424-8460(O)
2251-2608(P)

5. Epilogue

Generative AI provides an effective solution to address the traditional challenges in big data courses. Its innovative restructuring of teaching subjects, content, processes, and evaluations has been proven through practice to align with industry demands and skill development. Going forward, continuous optimization of human-machine collaboration mechanisms is essential to further integrate technology with pedagogy.

About the author

Hao Sun (1989-), born in Yancheng, Jiangsu, holds a bachelor's degree and is a senior engineer. His research interests include big data and cloud computing.

References

[1] Wang P . Research on the Transformation and Reconstruction of Modern Teaching Mode under the Background of Artificial Intelligence[J].Frontiers in Educational Research,2025,8(10).

[2] Hu Z ,Wu Y ,Chen S , et al. Research on the Practice of Reconstructing Classroom Teaching Mode Based on AIGC[J].Innovative Applications of AI,2025,2(2):122-131.

[3] Li Yuxian. Generative AI Empowers the Restructuring of Scientific PBL Teaching Model: An Integration Perspective on Resources, Process, and Evaluation [J]. Information Technology Education in Primary and Secondary Schools, 2025, (04):41-44.

[4] Akpan J I ,Razavi R ,Akpan A A .Evolutionary Trends in Decision Sciences Education Research from Simulation and Games to Big Data Analytics and Generative Artificial Intelligence.[J].Big data,2025.