

An AI Teaching Assistant for Blended Learning in a Big Data Course

Xiaojiao Guo, Zhanpeng Wen

School of Modern Information Industry, Guangzhou College of Commerce, Guangzhou, Guangdong, 511363, China

ABSTRACT

Online–offline blended learning has become a mainstream direction of higher-education reform, yet in courses that demand strong logical reasoning and programming skills—such as Spark Big Data Architecture—existing smart teaching tools still suffer from shallow use of learning data, weak real-time and personalized feedback, imprecise assessment of learning states, and a heavy analytical burden on teachers. To address these problems, this study designs and implements a blended teaching model supported by an AI teaching assistant (AI TA) and a course knowledge graph. Guided by course objectives, the model couples online autonomous learning with offline classroom instruction under a "teacher-led, student-centered, AI-assisted" framework. A Spark knowledge graph structures fragmented resources into a coherent knowledge network, while the AI TA—built on a large language model and a curated course knowledge base—provides round-the-clock question answering, source-grounded explanations, resource recommendation, and personalized support across the pre-class, in-class, and post-class stages. The model was applied in an undergraduate Spark course and evaluated through questionnaires and semi-structured interviews. Results show high student satisfaction: 92% strongly agreed with the overall model, 100% found it superior to traditional teaching for mastering content, and all respondents were willing to keep using it. The findings indicate that pairing a knowledge graph with an AI TA can effectively lower the comprehension threshold for abstract concepts, improve practice efficiency, and strengthen engagement, while the AI TA is best positioned as a support tool rather than a replacement for the teacher.

KEYWORDS

AI teaching assistant; Blended learning; Knowledge graph

1 Introduction

Talent cultivation is fundamental to social development, and education is the essential pathway to it. The application of artificial intelligence (AI)—and increasingly of large language models (LLMs)—in education can markedly improve teaching effectiveness, reduce teachers' workload, and provide adaptive learning support that helps cultivate new types of talent^[1]. As a prominent educational technology, an AI teaching assistant (AI TA) uses intelligent techniques to assist the design and delivery of instruction and to create a more efficient learning environment, and empirical studies report gains in student learning and engagement^[2]. More than a smart tool, an AI TA supports the transformation of higher education toward the vision of Education 4.0^[3], optimizing the teaching process, methods, and learning experience. Its core value lies in enhancing students' thinking: helping them acquire knowledge while developing reasoning, innovation, and problem-solving abilities^[1]. With the advance of information technology, online–offline blended learning has become the mainstream direction of teaching reform. Spark Big Data Architecture places equal emphasis on theory and practice and imposes high demands on students' logical thinking, programming, and data-processing skills. Building rich online resources and using smart tools to manage and assess the learning process can cultivate autonomous-learning habits and higher-order abilities. Nevertheless, four problems persist.

(1) Insufficient data mining. Although smart tools collect abundant behavioral data—study time, assignment submissions, online discussions—analysis often remains at a simple statistical level and fails to reveal deeper learning patterns.

(2) Limited timeliness and personalization. Learning trajectories are recorded, but immediate evaluation and personalized feedback are weak, and results often lag behind actual behavior, so students are not guided at critical moments.

(3) Imprecise assessment of learning states. Blended teaching can capture feedback, but assessment rarely progresses from individual points to the whole knowledge structure, making it hard to grasp learning states precisely in concept-heavy courses.

(4) Limited data-analysis capacity of teachers. Teachers' abilities to use tools and analyze data vary, and the heavy analytical workload constrains fine-grained management of the learning process.

In 2012, Google proposed the "knowledge graph," originally a knowledge base that organizes web data semantically to support intelligent search and reveal complex relations among entities. In recent years, scholars have introduced knowledge graphs into education to describe domain knowledge and its relations, helping students master a course's knowledge system, offering personalized learning paths, and helping teachers understand learning states. Supported by a domain knowledge graph, an adaptive learning system can accurately perceive a learner's state, progress, and weaknesses, understand real needs, and trigger adaptive and instructional interventions. Motivated by these opportunities, this study builds a blended teaching model that integrates a Spark knowledge graph with an AI TA and

evaluates it in an authentic course setting.

2 Related Work

An AI TA is an expert system developed from AI and cognitive theories to simulate the teaching process of a human tutor and provide efficient learning support. By modeling the interaction between student and tutor, it generates automated hints and feedback from a knowledge base and offers personalized, adaptive, fine-grained support tailored to a student's knowledge level and learning style. As large-model and big-data techniques mature, AI in education has attracted broad attention. Shi et al.^[1] and Cibu et al.^[4] systematically reviewed the educational applications of large language models and mapped their main use cases across teaching, learning, and assessment. Ivanova and Terzieva^[5] further surveyed how large language models are integrated into intelligent education systems, highlighting new opportunities for personalized, adaptive support as well as open challenges in reliability and pedagogical alignment.

Knowledge graphs have also been widely applied in education with notable value. Li et al.^[6] reviewed knowledge-graph-based personalized learning-path recommendation and showed that structuring domain knowledge as a graph enables systems to identify learners' states and suggest individualized paths. Lin and Wu^[7] combined a knowledge graph with knowledge tracing and a large language model to recommend adaptive learning paths, further strengthening personalization. These studies confirm that pairing a knowledge graph with intelligent, LLM-based tutoring is a promising route to personalized learning, which motivates the design in this paper.

3 The Proposed Blended Teaching Model

3.1 Overall Framework

The proposed model is guided by the teaching objectives of Spark Big Data Architecture and integrates online autonomous learning with offline classroom instruction to form a "teacher-led, student-centered, AI-assisted" framework. The teacher sets objectives, organizes content, and designs activities; students actively participate in knowledge construction and hands-on practice as the learning subject; and the AI TA serves as an intelligent support tool throughout, providing continuous knowledge services and guidance. Their coordination shifts instruction from one-way transmission toward multi-directional interaction and drives the model toward intelligence and personalization.

At the level of content organization, teaching materials are structured around a Spark knowledge graph that clarifies the logical relations and learning paths among knowledge points. Drawing on the graph and the course resource base, the AI TA provides precise knowledge localization and resource recommendation, enabling students to form a clear holistic picture of the complex Spark architecture. Online learning emphasizes knowledge acquisition and self-study, whereas offline learning focuses on explaining key and difficult points, discussion, and experimentation; the two complement each other to build an adaptive and sustainable blended model.

3.2 Functional Design of the AI TA

Within the model, the AI TA is positioned as a key support tool whose functions center on the needs of the Spark course. Grounded in the course knowledge base and knowledge graph, it can intelligently parse and answer questions about Spark core concepts, architectural principles, programming models, and experimental operations, offering timely and accurate support. During question answering it also provides the corresponding knowledge sources, strengthening the credibility of the content and helping students form systematic understanding.

Beyond passively answering questions, the AI TA proactively recommends resources and learning paths based on students' progress and interaction behavior. When a student repeatedly asks about a point or shows difficulty, it automatically pushes related videos, cases, or experiment guides to consolidate foundations and extend cognitive depth. By continuously collecting question data and feedback, it optimizes matching and response logic and gradually improves its ability to handle complex and diverse needs, thereby playing a stable, efficient supporting role.

3.3 Teaching Process Design

The process is designed around three connected stages—pre-class, in-class, and post-class—emphasizing continuity of activities and support.

Before class, the teacher designs tasks from the knowledge graph and objectives and releases requirements and resources via the course platform. While previewing, students use the AI TA to query points and resolve questions, forming an initial overview and lowering the entry barrier so that self-study becomes more efficient.

During class, instruction combines lecturing with participation and focuses on Spark architecture principles, core-component functions, and typical applications. Through case analysis and experiment demonstrations, the teacher guides students to understand real-world value, while the AI TA provides instant assistance—quick concept explanations and operation hints during discussion and practice—reducing barriers caused by individual differences and enhancing interactivity and relevance.

After class, the process emphasizes consolidation and improvement. While completing assignments and experiments, students continue to use the AI TA for troubleshooting and review. Based on behavioral data, it analyzes each student's situation and recommends review or extension resources for weak areas, deepening understanding of the Spark

knowledge system. Meanwhile, the teacher analyzes learning data and common problems through the platform to inform subsequent improvement, forming a data-driven teaching loop.

4 Implementation

4.1 Learner-characteristics Survey

To accurately grasp students' foundations and needs, a learner-characteristics questionnaire was administered at the start of the course, yielding 29 valid responses covering programming background, motivation, time investment, difficulties, and attitudes toward AI-assisted teaching. As summarized in Table 1, students generally possessed the basic programming foundation for the course, with wide exposure to Java, Python, and SQL, but showed clear gaps in Scala proficiency and understanding of distributed computing. They chose the course mainly out of interest in big-data technology and the practical need to improve employability, yet faced challenges bridging theory and practice due to limited time and experience. Most relied on teacher guidance, peer exchange, and AI tools, holding positive attitudes toward AI for explaining complex concepts, debugging code, and providing examples while remaining attentive to usage boundaries and information accuracy. Overall, learning exhibited "clear differences in foundations, strong practical needs, and insufficient personalized support," providing a realistic basis for combining a knowledge graph with an AI TA.

Table 1 Key results of the pre-course learner-characteristics survey

N = 29

Dimension	Representative items	Proportion (N = 29)
Programming background	Java / Python / SQL / Scala / R	93.1 / 93.1 / 82.8 / 51.7 / 17.2 %
Learning motivation	Interest in big data / Employability / Course credit	62.1 / 55.2 / 58.6 %
Weekly time investment	Ample / Average / Tight	44.8 / 44.8 / 10.3 %
Main challenges	Scala/Java syntax / Scarce resources / Distributed computing	65.5 / 65.5 / 62.1 %
Preferred help-seeking	Ask teacher / Ask peers / Use AI tools / Official docs	79.3 / 69.0 / 65.5 / 55.2 %
Most-wanted support	Shared practice guides / Detailed tutorials / More code examples	79.3 / 72.4 / 55.2 %

4.2 Building the Spark Knowledge Graph

By systematically restructuring the course content, key knowledge points were extracted and integrated with videos, test items, cases, and experiment guides, transforming fragmented resources into a structured knowledge network. Core points—Spark components (Spark Core, Spark SQL, Spark Streaming), the RDD programming model, data-processing workflows, optimization techniques, and typical scenarios—were organized to clarify their internal logic and build a hierarchy-centered graph framework. Each point was precisely linked to learning resources, forming a "knowledge-point profile" that includes an introduction plus related cases, experiment guidance, and extensions, so students can locate content quickly and move from single-point learning to holistic understanding. A learning-path recommendation function analyzes progress, weak areas, and interests to suggest optimal paths; for example, after finishing fundamentals, students are automatically pushed matching advanced content. Presented as nodes and links, the graph intuitively conveys dependencies and extension paths, giving students a holistic cognitive framework.

The graph was constructed bottom-up: entities and relations were first extracted from multiple sources—syllabi, textbooks, course videos, slides, and digital texts—to lay the entity layer, upon which the schema layer was progressively built and refined. To identify entities, the term frequency-inverse document frequency (TF-IDF) technique was applied to extract key concepts such as "RDD," "data partition," and "DAG scheduler." Extracted entities were then reviewed and consolidated to remove inaccurate titles and duplicates, and relations among entities were classified and defined as an ontology-based structured framework.

4.3 Building the AI TA

The AI TA was embedded in the course platform and connected to the document question-answering capability of a large language model to rapidly expand knowledge coverage, providing question answering, resource recommendation, intelligent item generation, and learning reminders, with an independent management console. Course materials—lecture notes, videos, cases, experiment manuals, and FAQs—were first uploaded to the knowledge base, and standard answers with similar phrasings were pre-entered for typical questions to ensure fast, accurate responses.

Leveraging deep learning over course materials, the AI TA supplies source information during question answering to strengthen credibility. Through the semantic-analysis and self-learning capabilities of the large model, it continuously improves its matching rate and satisfaction: by understanding the core of a question, it efficiently matches standard answers, and based on satisfaction feedback it dynamically adjusts knowledge-base content and matching logic, focusing on weak question types to gradually handle complex or ambiguous queries. The AI TA offers 24/7 uninterrupted question answering around core content, delivering clear, well-organized answers for both basic understanding and complex discussion and reducing the frustration caused by learning obstacles.

5 Evaluation

To systematically assess the model's effect in the Spark course, questionnaires and interviews were combined to

evaluate satisfaction, learning support, engagement, and acceptance.

5.1 Questionnaire Results

After the course, a five-point Likert-scale satisfaction survey of the AI-TA blended teaching model was distributed, comprising seven core items on overall satisfaction, AI-TA function, efficiency gains, engagement, and continued-use intention; 25 valid responses were collected. As shown in Table 2, evaluations were highly positive. For overall satisfaction, 92% strongly agreed and 8% agreed. For learning support, 100% agreed that the AI TA responded in a timely and effective way, with 88% strongly agreeing. For understanding abstract concepts such as Spark architecture and RDD principles, 88% strongly agreed and 12% agreed. For code understanding, debugging, and experiments, 84% reported markedly improved efficiency. For engagement, 96% felt greater participation and initiative. Notably, 100% considered the “AI TA + blended” model more helpful than traditional teaching, and all respondents were willing to keep using it—evidence of strong recognition and promotion potential.

Table 2 Results of the post-course satisfaction survey

N = 25

#	Survey item	Strongly agree	Agree
1	Overall satisfaction with the AI-TA blended model	92%	8%
2	The AI TA responds to problems in a timely, effective way	88%	12%
3	The AI TA helps explain abstract concepts (Spark architecture, RDD)	88%	12%
4	The AI TA supports code understanding, debugging and experiments	84%	16%
5	The model enhanced my engagement and learning initiative	96%	4%
6	More helpful than traditional teaching for mastering content	100%	0%
7	Willing to keep using the model in future related courses	100%	0%

5.2 Interview Results

Semi-structured interviews with selected students indicated that the AI TA effectively compensated for insufficient after-class support, especially during experiment debugging and knowledge review, where targeted guidance lowered difficulty. Its “ask-anytime, answer-anytime” nature gave students a greater sense of security and reduced frustration from accumulated problems. Some students noted that the AI TA suits basic understanding and operational guidance, whereas comprehensive problems and project design still require teacher guidance and classroom discussion. This suggests the AI TA should be positioned as an “assistive supporter” rather than a replacement for the teacher, which aligns closely with the philosophy of blended learning.

6 Discussion and Conclusion

Combining questionnaire and interview results, the AI TA clearly promoted learning outcomes in the Spark course. On one hand, timely question answering and diverse explanations lowered the threshold for understanding abstract concepts and complex techniques; on the other, support for code examples, debugging advice, and experiment guidance improved practical efficiency and problem-solving. Continuous, companion-style support also strengthened confidence and the willingness to learn autonomously, helping cultivate good learning habits.

Overall, the AI-TA-based blended model offers clear advantages in raising satisfaction, enhancing engagement, and supporting personalized learning. It organically integrates classroom teaching, online learning, and intelligent assistance, preserving the teacher’s leading role in design and value guidance while leveraging AI to broaden and deepen support. The study also finds that reasonable use of AI tools must be guided to avoid over-reliance, and that the AI TA remains limited for complex, comprehensive problems and higher-order ability development, which should be complemented by teacher guidance, group discussion, and project practice. Future work will optimize AI-TA functions and further explore a long-term mechanism for coordinated development between the AI TA and the teacher.

About the Author

Xiaojiao Guo, master's degree, teaching assistant, research directions are big data technology and artificial intelligence-assisted teaching.

References

- [1] SHI Y, YU K, DONG Y, et al. Large language models in education: a systematic review of empirical applications, benefits, and challenges[J]. Computers and Education: Artificial Intelligence, 2026, 10: 100529.
- [2] ROCA M D L, CHAN M M, GARCIA-CABOT A, et al. The impact of a chatbot working as an assistant in a course for supporting student learning and engagement[J]. Computer Applications in Engineering Education, 2024, 32(5): e22750.
- [3] PELÁEZ-SÁNCHEZ I C, VELARDE-CAMAQUI D, GLASSERMAN-MORALES L D. The impact of large language models on higher education: exploring the connection between AI and Education 4.0[J]. Frontiers in Education, 2024, 9: 1392091.
- [4] CIBU B R, CRĂCIUN L, MOLĂNESCU A G, et al. Exploring the educational applications of large language models: a systematic review and topic analysis[J]. Electronics, 2025, 14(23): 4683.
- [5] IVANOVA T, TERZIEVA V. Large language models in intelligent education systems: new educational perspectives—a systematic review[J]. Information, 2026, 17(5): 433.
- [6] LI A, LI Y, GAO X. Personalized learning path recommendation based on knowledge graphs: a survey[J]. Electronics, 2026, 15(1): 238.
- [7] LIN Y, WU Z. Learning path recommendation enhanced by knowledge tracing and large language model[J]. Electronics, 2025, 14(22): 4385.