2025 International Conference on Economics, Management and Education Technology (ICEMET 2025)

Article

Optimization and Application of Large Language Model in Higher Education Question Answering System

Hui Jia [1,*] and Kaimei Luo [2]

[1] Chinese Language Center, Shenzhen MSU-BIT University, Shenzhen, China
[2] Zhongshan Huachen Experimental Middle School, Zhongshan, China
* Correspondence: Hui Jia, Chinese Language Center, Shenzhen MSU-BIT University, Shenzhen, China

Abstract: With the rapid development of large language model technology, its application in diverse domains has expanded significantly, and higher education is no exception. This study investigates the optimization and application of large language models in higher education knowledge question-answering systems. First, it reviews the theoretical foundations of large language models and domain-oriented question-answering systems, clarifying their core architectures, training paradigms, and typical deployment modes in educational contexts. Second, it examines strategies to enhance system performance from three key dimensions: data and model optimization, integration of structured and unstructured knowledge resources, and the design of effective human–computer interaction mechanisms. On this basis, the study elaborates on innovative applications across disciplines, including intelligent tutoring, personalized learning support, formative assessment, academic research assistance, and institutional knowledge management. Furthermore, it analyzes major challenges such as data quality, model bias, interpretability, privacy protection, and alignment with pedagogical goals, and summarizes corresponding coping strategies and governance frameworks. Finally, the paper discusses future development trends, including multimodal integration, adaptive learning analytics, and closer collaboration between educators and AI systems. The study aims to provide a systematic reference for the effective deployment of large language models in higher education knowledge question-answering systems and to promote the transformation and sustainable development of higher education teaching models.

Keywords: large language models; higher education; question answering; knowledge management; teaching innovation; educational technology

Received: 06 February 2026
Revised: 25 March 2026
Accepted: 04 April 2026
Published: 11 April 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Large language model technology has achieved remarkable advancements in the domain of natural language processing, with continuous improvements in model architecture and training methodologies. Despite these advancements, existing higher education knowledge question-answering systems face notable challenges in addressing complex queries, ensuring timely updates to knowledge bases, and delivering personalized services tailored to individual needs. The integration of large language models into these systems presents substantial potential for innovation and enhancement. This integration can drive transformative changes in higher education teaching methodologies, shifting away from traditional knowledge dissemination approaches toward more interactive and personalized learning experiences. Furthermore, it can significantly enhance students' learning outcomes by enabling them to acquire knowledge and solve problems with greater efficiency. Additionally, this approach can contribute to the equitable distribution of educational resources, mitigating disparities caused by regional and institutional differences. By leveraging advanced technologies, more students can access high-quality educational materials and opportunities. This study

employs a combination of literature analysis, case studies, empirical investigations, and other research methods to thoroughly explore the optimization and application of large language models within higher education knowledge question-answering systems.

2. Theoretical Foundation: Analysis of Large Language Model and Higher Education Knowledge Question Answering System

2.1. Large Language Model: Its Concept and Development History

A Large Language Model (LLM) is an advanced artificial intelligence system designed to comprehend, generate, and manipulate human language using deep learning techniques. These models are built upon transformer architectures, which utilize neural networks with self-attention mechanisms to process and analyze language data effectively. They are trained on vast and diverse datasets, including text from books, websites, articles, and other linguistic sources, enabling them to identify patterns, semantics, syntax, and contextual relationships within language. This extensive training allows LLMs to understand context and generate coherent, human-like text. With billions or even trillions of adjustable parameters, these models optimize their ability to map inputs to outputs, making them highly versatile for performing a wide range of natural language processing tasks [1]. These tasks include text generation, summarization, translation, question answering, and sentiment analysis, all achieved without requiring task-specific retraining. The adaptability and scalability of LLMs have positioned them as transformative tools in various fields, including higher education, where they can enhance learning experiences and streamline academic processes.

Among the foundational technical principles of large language models, the Transformer architecture plays a pivotal role in capturing long-distance dependencies within text through its self-attention mechanism. This significantly enhances the model's ability to understand complex language structures. Self-supervised learning further empowers these models by enabling them to autonomously discover patterns within large volumes of unlabeled data, thereby reducing reliance on manually labeled datasets. For instance, generative pre-training methods allow models to produce coherent and contextually relevant text, while bidirectional encoding techniques excel in semantic comprehension. These advancements have led to the development of models capable of performing diverse tasks with remarkable efficiency. The evolution of large language models reflects continuous improvements in their understanding and reasoning capabilities, driven by the increasing number of parameters and advancements in training methodologies. As these models become more compact and practical, they are increasingly accessible for applications in higher education, offering innovative solutions for academic research, teaching, and administrative tasks [2].

In earlier models such as Word2Vec, each word was associated with a fixed meaning, limiting their ability to adapt to contextual variations. Subsequent advancements, such as the introduction of bidirectional encoding in newer models, have enabled a more nuanced understanding of words based on their surrounding context. While models like BERT excel in reading comprehension and semantic analysis, they do not actively generate content [3]. In contrast, the GPT series of models has demonstrated exceptional capabilities in generating text and addressing open-ended questions. These models follow a methodology of extensive pre-training followed by task-specific fine-tuning, which enhances their performance across various applications. The progression from simpler models to today's sophisticated systems has been marked by significant improvements in both functionality and accessibility. For example, newer iterations like GPT-4 exhibit enhanced reasoning and understanding capabilities, while technological advancements have made it possible to run these models on standard computing devices. Looking ahead, large language models are expected to achieve breakthroughs in areas such as multimodal fusion and model compression, further broadening their applicability and impact across diverse domains, including higher education.

2.2. Higher Education Knowledge Question Answering System: Current Situation and Problems

Higher education knowledge question-and-answer systems typically involve several key processes, including knowledge acquisition, storage, retrieval, and answer generation. During the knowledge acquisition phase, relevant information is gathered from diverse educational resources, ensuring a broad and comprehensive knowledge base. The storage phase organizes this information into formats that facilitate efficient retrieval, while the retrieval phase focuses on identifying pertinent data based on user queries. Finally, the answer generation phase synthesizes and processes the retrieved information to deliver coherent and meaningful responses to users. These systems have been implemented in various scenarios, such as course-related question-and-answer sessions, academic consultations, and exam preparation assistance [4, 5]. While they offer some benefits to students, user feedback highlights significant shortcomings, particularly in terms of knowledge accuracy. Many systems frequently provide incorrect or overly vague answers, undermining their reliability. Additionally, their reasoning capabilities are often inadequate, especially when addressing complex questions that demand logical analysis. The user interaction experience also requires substantial improvement, as the current interaction processes are often perceived as unnatural and inconvenient, limiting their overall effectiveness.

In practical applications, several critical issues have been identified. One major problem is the slow pace of knowledge updates, which typically occur only once every one or two years. This delay prevents systems from reflecting the latest advancements in rapidly evolving disciplines [6]. For instance, in the field of computer science, when students inquire about comparisons such as "Which is better, Transformer or RNN?", some systems provide outdated responses that fail to account for recent developments. Another significant limitation is the lack of complex reasoning capabilities. When faced with multi-step analytical questions, such as "Using economic principles to analyze the impact of inflation on unemployment rate," these systems often present fragmented knowledge points without constructing a coherent logical explanation. Furthermore, the interaction design of these systems is overly simplistic, making it difficult for users to engage in seamless and intuitive exchanges. These challenges primarily stem from the traditional systems' limited ability to comprehend and process language at a deeper level, which restricts their capacity to address the intricate demands of higher education. Addressing these issues requires advancements in language understanding and system design to better align with the sophisticated needs of academic users.

3. Optimization Method: Strategies for Improving the Performance of Question-Answering Systems Using Large Language Models

3.1. Optimization Measures for Data and Models

At the data level, constructing and expanding high-quality datasets serves as a fundamental step in optimizing question-answering systems. This involves collecting relevant text data from diverse sources, such as course materials, academic papers, and textbooks from educational institutions, and organizing and annotating them to ensure they meet specific requirements. Employing data cleaning techniques is essential to eliminate noise and redundant information, thereby improving the overall quality of the dataset [7, 8]. Pre-processing methods, such as annotation, further enhance the usability and relevance of the data. Additionally, data augmentation techniques, including text synonym replacement and sentence transformation, can be utilized to increase the diversity of the dataset, which in turn strengthens the model's ability to generalize across various contexts. At the model level, it is crucial to evaluate the suitability of different large language models based on the specific objectives of the system and select the most appropriate one. Experimental analysis should be conducted to fine-tune hyperparameters, identifying the optimal configuration that maximizes model performance. Furthermore, exploring advanced strategies such as model fusion and ensemble learning can leverage the strengths of multiple models, thereby achieving a

higher level of accuracy and efficiency in the system's performance. These measures collectively contribute to the development of robust and effective question-answering systems.

3.2. Methods at the Knowledge and Interaction Level

At the knowledge level, integrating knowledge graphs and large language models represents a significant optimization direction. Knowledge graphs provide a structured and systematic representation of information, which, when combined with large language models, can substantially enhance reasoning capabilities. By refining model architectures and training methodologies, the system's logical reasoning and semantic understanding can be further improved. For addressing complex problems, decomposition strategies are employed to break them into smaller, manageable sub-problems. This approach facilitates multi-step reasoning, thereby increasing the accuracy and reliability of the solutions generated. At the interaction level, the design and refinement of natural language interfaces are of paramount importance. These interfaces should be intuitive, user-friendly, and efficient, enabling seamless interaction between users and the system. Developing personalized answer generation and recommendation mechanisms is highly recommended. Such mechanisms can tailor responses and suggested content to align with individual user histories, interests, and preferences, thereby enhancing user engagement and satisfaction. Additionally, implementing a robust user feedback collection system is essential. This system should actively gather evaluations and suggestions from users, enabling iterative improvements to the system. By continuously optimizing based on user input, the system can achieve higher performance levels and ensure a more satisfying user experience.

4. Innovative Application of Large Language Model in Higher Education Knowledge Question Answering System

4.1. Analysis of Application Examples in Different Disciplines

In the fields of science and engineering, large language models play a significant role in answering professional questions and supporting scientific research. For example, in computer science, these models can greatly enhance automated answering systems for algorithm design courses. When students pose a question such as "how to optimize the shortest path algorithm based on dynamic programming ideas," the model can analyze the core elements of dynamic programming, including state definition, transfer equations, and boundary conditions. It can then identify the limitations of Dijkstra's algorithm and propose optimization strategies. During this process, the model can automatically associate the relaxation operation principle of the Bellman-Ford algorithm and generate pseudocode examples to aid comprehension. This capability transforms abstract algorithmic concepts into actionable implementation steps, effectively bridging the gap between theoretical understanding and practical application in traditional teaching.

In the field of liberal arts, the application of large language models in history education demonstrates their strength in text analysis. For instance, when addressing a question like "Comparing the institutional reform logic of two historical reforms," the model can construct a two-dimensional analytical framework. This framework might involve comparing the motivations for change and the underlying power structures. The model can then draw upon historical records to substantiate its analysis and extract broader patterns, such as the principle that the sustainability of institutional change often hinges on the mechanisms of interest distribution [9, 10]. By doing so, the model helps students develop a structured methodological approach to conducting comparative historical research, fostering deeper analytical skills and a more nuanced understanding of historical processes.

Applications in the medical field underscore the importance of professionalism and precision. For example, in anatomy education, when students inquire about "the pathological mechanism and clinical relevance of the hypothalamus-pituitary-adrenal

axis," the model can provide a detailed explanation of the secretion regulation pathway of corticotropin-releasing hormone (CRH). It can further elaborate on the classification of Cushing's syndrome into ACTH-dependent and ACTH-independent categories, offering insights into their respective etiologies. Additionally, the model can integrate clinical case studies to illustrate the diagnostic significance of imaging techniques and hormone level assessments. The core value of this application lies in its ability to connect fragmented physiological knowledge with the logical framework of clinical diagnosis and treatment, thereby creating a comprehensive cognitive chain that enhances both theoretical understanding and practical application (As shown in Figure 1).

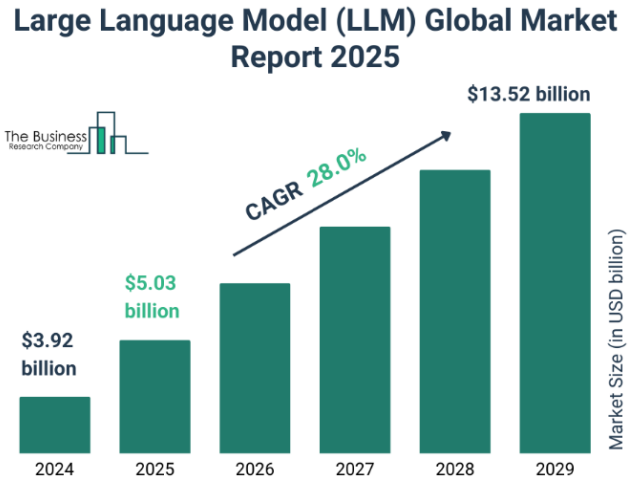

Figure 1. Large language model global market report 2025.

4.2. Multiple Application Modes in Teaching and Learning Scenarios

In classroom teaching, integrating the large language model into the teaching platform enables real-time question-answering and interactive support functions. Teachers can utilize this system to address students' questions promptly during class, thereby fostering a more engaging and participatory learning environment. In the context of after-school self-study, the large language model serves as an intelligent tutoring tool, offering personalized learning guidance tailored to individual students' needs. By analyzing students' learning progress and identifying areas of difficulty, the system can design customized learning strategies to enhance efficiency. During self-study sessions, the personalized learning paths generated by the large language model demonstrate significant advantages. For instance, after evaluating students' exercise responses and pinpointing knowledge gaps, the system can automatically recommend progressive learning resources. At the foundational level, it may provide instructional videos that combine theoretical concepts with practical examples, such as using probability calculations in medical diagnostics. At the advanced level, it can generate targeted exercises that transition from basic probability problems to more complex scenarios involving multi-conditional probability. Furthermore, at the extension level, the system can suggest interdisciplinary application materials, such as literature on naive Bayes classifiers in machine learning [11]. This precise resource recommendation mechanism effectively addresses the challenge of "resource overload with insufficient adaptability" that often hinders traditional self-study approaches.

For examinations and assessments, the innovative application of the large language model is evident in two key areas: dynamic test creation and intelligent review processes. During test creation, the model can automatically generate multi-tiered test papers aligned with the course outline, encompassing basic, comprehensive, and advanced questions. This ensures a balanced evaluation of students' understanding across different levels of complexity. In the review phase, the model employs semantic similarity analysis and logical chain evaluation to provide in-depth feedback on open-ended discussion

questions. For example, it can identify specific knowledge gaps, such as neglecting to consider cost-driven factors when analyzing economic phenomena like inflation. Additionally, it can offer constructive suggestions, such as emphasizing the need for data-supported arguments in responses. By automating these processes, the system significantly reduces the workload for educators while delivering more detailed and actionable feedback compared to traditional scoring methods. This dual functionality not only streamlines the assessment process but also enhances the overall quality of feedback provided to students, fostering a deeper understanding of the subject matter.

4.3. In-depth Application in Academic Research and Knowledge Management

This section explores the advanced application of large language models in academic research and knowledge management. These models play a pivotal role in enhancing the efficiency of literature retrieval and review generation by enabling researchers to quickly locate relevant studies, systematically organize findings, and produce comprehensive literature reviews. This significantly streamlines the research process and allows scholars to focus on deeper analytical tasks. Furthermore, in addressing complex scientific challenges and fostering innovation, large language models serve as a valuable tool by offering diverse perspectives and methodologies. They stimulate creative thinking and contribute to the development of groundbreaking ideas in various academic fields [1, 12]. Additionally, these models are instrumental in constructing and maintaining academic knowledge graphs. By collecting, structuring, and analyzing vast amounts of academic data, they enable more effective management and utilization of knowledge resources, thereby providing robust support for ongoing and future research endeavors.

5. Challenges and Prospects: Dilemmas and Future Trends in the Application of Large Language Models

5.1. Challenges and Coping Strategies in Application

On a technical level, large language models face significant challenges related to interpretability, as their "black box" nature makes it difficult for users to comprehend the underlying decision-making processes. Enhancing the interpretability of these models can be achieved through advanced techniques such as visualization tools and attribution analysis, which provide insights into how decisions are made. Furthermore, these models require substantial computing resources, which can be mitigated by adopting efficient computing solutions, including cloud computing, distributed computing frameworks, and model compression techniques. Another critical area for improvement is the generalization ability and domain adaptability of these models. By employing strategies such as multi-domain training and domain adaptation, their performance across diverse fields and tasks can be significantly enhanced, ensuring broader applicability and reliability.

From an ethical and security perspective, safeguarding data privacy is paramount. Establishing a comprehensive data management framework is essential to ensure the protection of user privacy throughout the processes of data collection, storage, and utilization. Addressing algorithmic bias and discrimination is another critical concern, which can be tackled through meticulous data preprocessing and algorithm refinement [13, 14]. Additionally, to prevent the generation of false information and uphold academic integrity, it is crucial to implement stringent review mechanisms and supervision protocols for the content produced by large language models. These measures will help maintain trust and reliability in the application of such technologies.

At the educational and teaching level, the integration of large language models has brought transformative changes to the roles and methodologies of educators [15]. Teachers are required to shift from being mere transmitters of knowledge to becoming facilitators and guides in the learning process. This transition necessitates the adoption of innovative teaching strategies that effectively incorporate large language models into educational practices. Concurrently, it is vital to provide students with proper guidance

on the responsible use of these models, fostering their critical thinking, creativity, and self-directed learning capabilities. Establishing a robust evaluation system tailored to the evolving educational landscape is equally important, ensuring that assessments align with the new paradigms introduced by these advanced technologies.

5.2. Future Development Trends

In the future, the application prospects of multimodal fusion technology in knowledge question and answer systems are extensive and promising. By integrating multimodal information, such as text, images, and audio, into large language models, these systems can achieve enhanced interactivity and improved accuracy, offering users access to more comprehensive and enriched information. The coordinated development of federated learning and privacy protection technologies is expected to further optimize model performance while safeguarding data privacy. This approach allows data to remain localized while still contributing to model training, ensuring a balance between innovation and security [2]. Additionally, advancements in quantum computing technology hold the potential to significantly accelerate the training processes and computational capabilities of large language models, thereby driving transformative changes in their applications across various domains. In the realm of education, the expansion and innovation of large language models are anticipated to play a pivotal role in lifelong learning and vocational education, offering tailored learning support to individuals across diverse age groups and professional requirements. These models also facilitate cross-cultural education and international exchanges, fostering knowledge sharing and communication across different cultural contexts. The integration of emerging educational technologies, such as virtual reality, augmented reality, and blockchain, is expected to create novel application scenarios, further promoting educational innovation and development. From a policy and ethical perspective, it is essential for governments and educational institutions to establish comprehensive guidelines to support the healthy development of large language models in higher education. Industry-wide self-regulation mechanisms and ethical standards should be implemented to ensure the responsible use of these technologies. Furthermore, international collaboration is crucial for developing unified standards and norms, which can facilitate the orderly and effective application of large language models in the field of higher education.

6. Conclusion

The optimization and application of large language models in higher education knowledge question-answering systems hold immense potential for transforming the educational landscape. These models have demonstrated innovative applications across diverse subject areas, enhancing teaching methodologies, enriching learning experiences, advancing academic research, and streamlining knowledge management processes. By doing so, they contribute significantly to improving educational quality and boosting the efficiency of scientific research. However, the integration of these models is not without challenges. Issues related to technological limitations, ethical considerations, data security, and the adaptation of educational practices must be addressed through well-thought-out strategies. As technology continues to evolve and regulatory frameworks become more robust, the scope for applying large language models in higher education is expected to expand, enabling deeper and more impactful integration. Nevertheless, this study acknowledges certain limitations, particularly in the comprehensiveness of the data analyzed and the verification of the practical effectiveness of these models. Future research should focus on addressing these gaps by exploring more comprehensive datasets and conducting rigorous evaluations of real-world applications. Such efforts will pave the way for innovative methodologies that further integrate large language models into higher education, fostering sustainable development and long-term advancements in this field.

References

1. V. Nikolovski, D. Trajanov, and I. Chorbev, "Advancing AI in higher education: A comparative study of large language model-based agents for exam question generation, improvement, and evaluation," *Algorithms*, vol. 18, no. 3, p. 144, 2025.
2. A. R. Swisher, A. W. Wu, G. C. Liu, M. K. Lee, T. R. Carle, and D. M. Tang, "Enhancing health literacy: evaluating the readability of patient handouts revised by ChatGPT's large language model," *Otolaryngology–Head and Neck Surgery*, vol. 171, no. 6, pp. 1751–1757, 2024.
3. M. Zhang, A. Pernille Dilling, L. Gondelman, N. E. Ruan Lyngdorf, E. D. Lindsay, and J. Bjerva, "SEFL: Harnessing large language model agents to improve educational feedback systems," *arXiv e-prints*, arXiv-2502, 2025.
4. A. K. Shanthakumar, F. Fassihi-Tash, A. Lotfi, and J. J. Bird, "Retrieval augmented large language model chatbots in higher education: A study on university open days," in *UK Workshop on Computational Intelligence*, Cham: Springer Nature Switzerland, pp. 32–44, Sept. 2024.
5. A. Al-Dokhny, O. Alismaiel, S. Youssif, N. Nasr, A. Drwish, and A. Samir, "Can multimodal large language models enhance performance benefits among higher education students? An investigation based on the task–technology fit theory and the artificial intelligence device use acceptance model," *Sustainability*, vol. 16, no. 23, p. 10780, 2024.
6. O. Tayan, A. Hassan, K. Khankan, and S. Askool, "Considerations for adapting higher education technology courses for AI large language models: A critical review of the impact of ChatGPT," *Machine Learning with Applications*, vol. 15, p. 100513, 2024.
7. I. C. Peláez-Sánchez, D. Velarde-Camaqui, and L. D. Glasserman-Morales, "The impact of large language models on higher education: exploring the connection between AI and Education 4.0," in *Frontiers in Education*, vol. 9, p. 1392091, June 2024.
8. T. Alqahtani, H. A. Badreldin, M. Alrashed, A. I. Alshaya, S. S. Alghamdi, K. Bin Saleh, and A. M. Albekairy, "The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research," *Research in Social and Administrative Pharmacy*, vol. 19, no. 8, pp. 1236–1242, 2023.
9. J. Savelka, A. Agarwal, M. An, C. Bogart, and M. Sakr, "Thrilled by your progress! Large language models (GPT-4) no longer struggle to pass assessments in higher education programming courses," in *Proceedings of the 2023 ACM Conference on International Computing Education Research-Volume 1*, pp. 78–92, Aug. 2023.
10. S. Milano, J. A. McGrane, and S. Leonelli, "Large language models challenge the future of higher education," *Nature Machine Intelligence*, vol. 5, no. 4, pp. 333–334, 2023.
11. E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, and G. Kasneci, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.
12. S. Laato, B. Morschheuser, J. Hamari, and J. Björne, "AI-assisted learning with ChatGPT and large language models: Implications for higher education," in *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*, pp. 226–230, July 2023.
13. M. D. Idris, X. Feng, and V. Dyo, "Revolutionizing higher education: Unleashing the potential of large language models for strategic transformation," *IEEE Access*, vol. 12, pp. 67738–67757, 2024.
14. J. G. Meyer, R. J. Urbanowicz, P. C. Martin, K. O'Connor, R. Li, P. C. Peng, and J. H. Moore, "ChatGPT and large language models in academia: opportunities and challenges," *BioData Mining*, vol. 16, no. 1, p. 20, 2023.
15. D. Yigci, M. Eryilmaz, A. K. Yetisen, S. Tasoglu, and A. Ozcan, "Large language model-based chatbots in higher education," *Advanced Intelligent Systems*, vol. 7, no. 3, p. 2400429, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.