https://doi.org/10.37420/j.cer.2026.091

Multimodal AI Teaching Assistants: A Study on the New Paradigm for College English Teaching

Jing Tian*
Guangdong Baiyun University, Guangzhou, China
***Corresponding author, E-mail: oldcat6702024@163.com**

Abstract

This article explores the way of optimizing conventional teaching paradigms by tracing the development of the integration of artificial intelligence (AI) with multimodal teaching of College English. The study argues that by adhering to a tripartite symbiotic model of "teacher–Multimodal AI Teaching Assistant–student", AI applications and generative large models could act as teaching assistants to improve the multimodal College English teaching process.

Keywords:

Multimodal teaching, Digital transformation, AI empowerment

1 Introduction

Artificial intelligence is a major driving force behind the current wave of scientific and technological revolution and industrial transformation. Against this background, the integration of AI technologies with multimodal teaching has become a prominent topic in the latest round of educational reform.

College English is a compulsory general education course for non-English majors in Chinese universities. It aims to improve students' comprehensive English abilities, including listening, speaking, reading and writing, which is significant in cultivating their intercultural communication competence.

Nevertheless, current reforms in College English teaching remain caught in a practical dilemma of "high input but low effectiveness." On the one hand, although AI has gradually entered College English classrooms, most applications remain confined to isolated functions such as automated essay correction and have yet to connect the integrated chain of teaching, learning, assessment, and management. The overall potential of technological empowerment has therefore not been fully realized. Despite the considerable effort that many teachers devote to instructional design, limited access to information about students' learning makes it difficult to adjust teaching dynamically on the basis of comprehensive, real-time data. On the other hand, English proficiency integrates listening, speaking, reading, writing, and translation, and conventional teaching models have achieved only limited success in developing these abilities comprehensively. A growing range of AI-enabled educational applications can now integrate textual, auditory, visual, behavioral, and affective data, providing a practical technological foundation for the deep incorporation of multimodal AI tools into College English reform.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

2 Review of research on the integration of multimodal teaching paradigms and ai technologies

Traditional College English teaching is single-mode with text-centered instruction. With the advancement of digital technology, especially the emergence of AI technologies, College English teaching is undergoing a paradigmatic shift. Since AI can "generate multichannel stimuli and reduce the cognitive load of language learning," concretize topical contexts, and help produce "creative and imaginative audio or visual resources conducive to cultivating learners' imagination and innovative thinking", AI-empowered multimodal teaching

With the rapid evolution of multimedia, big-data, and AI technologies, : from single-mode, text-centered instruction to multimodal integration, and from conventional lecture-based classrooms to intelligent, interactive learning environments. AI can "generate multichannel stimuli and reduce the cognitive load of language learning," concretize topical contexts, and help produce "creative and imaginative audio or visual resources conducive to cultivating learners' imagination and innovative thinking," thereby substantially improving instructional effectiveness (Ou et al., 2024).

Multimodality refers to the use of multiple semiotic systems or communicative channels—including language, images, sound, gesture, and facial expression—to communicate and convey information. Multimodal theory can be traced back to multimodal discourse analysis, which is rooted in Hallidayan social semiotics. Building on this theoretical foundation, Kress and van Leeuwen systematically developed a theory of visual grammar in their seminal work, Reading Images: The Grammar of Visual Design, explaining how different semiotic resources work together to construct representational, interactive, and compositional meanings. Their work marked the formal establishment of multimodal discourse analysis as a discipline (Kress & van Leeuwen, 2006). In 2001, they further expanded the scope of semiotic research, arguing that meaning is not produced solely through linguistic text: images, sound, color, and action can also function as social semiotic resources (i.e., "modes"), while the material and technological carriers that bear these modes constitute "media." Therefore, different modes rely on media to jointly construct communicative meaning (Kress & van Leeuwen, 2001). Recent statistical evidence suggests that multimodal theory became a prominent topic in educational research both in China and internationally around 2010. After 2015, the widespread adoption of mobile internet technologies and smart devices further enriched multimodal teaching through video instruction, interactive courseware, virtual simulation, and other means. Since 2022, multimodal teaching has entered a new phase of deep integration with artificial intelligence. However, existing research tends to focus on applications within individual courses or specific disciplinary contexts, emphasizing scattered examples of application rather than systematically reviewing and integrating the field's overall developmental trajectory and cutting-edge developments.(Wang, Z., & Chen, 2026).

In Chinese education research, multimodal teaching research began with the exploration of innovative models and principles of multimodal teaching. In 2019, Yang Ping proposed to use "big data" and "artificial intelligence" technology to improve the traditional writing evaluation system. Under the guidance of a multimodal network platform, process writing theory and self-efficacy theory, automatic online correction tools were introduced to construct an interactive evaluation model of teacher-student dual subjects and a multimodal teaching model of human-computer collaboration (Yang, 2019). In 2010, Zhang Delu systematically

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

analyzed the principles of modal selection in different instructional designs : optimal principle, effective principle, adaptation principle and economic principle (Zhang, D., 2010). These principles and the extended sub-principles provide a reference framework for teachers to apply multimodal teaching methods in teaching research. Since then, multimodal teaching research has not only continued to explore the macro path, but also extended to the cultivation of English listening, speaking, reading, writing, translation and other detailed abilities.

After that, the research on multimodal teaching paradigm continues to extend to the macro path of artificial intelligence-enabled multimodal teaching. The existing research points out the dilemma of multimodal English teaching assisted by AI and puts forward different solutions from different perspectives, such as paying attention to teachers or students or paying attention to technology development. Jia Huihui (2022) analyzed the dilemma of multimodal teaching of college English from four dimensions : culture, context, content and expression, and puts forward innovative paths such as developing multimodal context teaching with the help of modern information technology and adhering to the unity of meaning and form (Jia, 2022). From the perspective of educational innovation, Ma Xiaowei (2024) called for the implementation of multimodal dialogue teaching in combination with the information environment and English teaching practice to improve the quality of teaching (Ma, 2024). Huo Nan (2025) focused on the solution of the problem of " insufficient participation " in traditional classrooms by multi-modal intelligent interaction technology, and proposes to pay attention to the role transformation of teachers and help students build adaptive personalized learning paths (Huo, 2025). Overall, these studies are mostly theoretical discussions or experience summaries, which provide macro ideas for follow-up research, but generally lack systematic data verification.

In terms of online teaching, reading teaching and classroom interaction, AI-assisted multimodal teaching expands multimodal teaching from traditional classrooms to specific scenarios such as online, reading and interactive design. Although some innovative multimodal teaching models are proposed, empirical verification is weak. Aiming at the problems of insufficient technology empowerment, lagging content, insufficient participation and single evaluation in college English online teaching under the multimodal interactive environment, Liu (2025) proposed comprehensive optimization strategies such as technology empowerment, situation construction, interaction enhancement and evaluation innovation (Liu, S., 2025). Zhou Chunhui (2025) explored an AI-supported model of multimodal reading instruction and argued that such a model should strengthen students' multimodal information-processing abilities and intercultural reading literacy (Zhou, 2025). Alitunur (2026) used intelligent question-answering systems, speech recognition, and visual recognition to construct a collaborative "teacher–student–machine" interaction model intended to address insufficient teacher–student interaction and unmet personalized learning needs in large classes (Alitunur, 2026).

In writing instruction, the integration of multimodal teaching and AI has progressed from advocacy of new models to the reconstruction and empirical examination of assessment systems. Using questionnaires together with quantitative and qualitative analyses, Wang Yilei (2024) evaluated the effectiveness of an AI-enabled multimodal College English writing workshop. The study affirmed a multimodal model combining AI-assisted learning with AI-supported teacher and peer assessment, while noting persistent time-management difficulties in timed writing tasks (Wang, 2024).

In summary, existing studies widely acknowledge the positive role and importance of multimodality and AI

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

in transforming College English teaching and emphasize the reconstruction of teaching models through the coordinated participation of teachers, students, and AI technologies. Most studies, however, remain focused on broad theoretical discussions and proposed pathways. Further research is needed on integrating multimodal instructional design with AI, establishing sound procedures for pre-class design and classroom implementation, and developing more fine-grained systems of feedback and assessment.

In recent years, an important application of the integration of multimodal instructional design and AI technology is “multimodal AI teaching assistant”. Multimodal AI teaching assistant is the product of the deep integration of generative artificial intelligence technology and the education field, which is promoting the paradigm change in the field of education. Its core advantage is its powerful content generation and understanding ability, which can efficiently and accurately respond to complex teaching scenarios quickly. It can not only intelligently integrate teachers ' multimodal teaching resources, but also customize personalized multimodal learning paths for students (Lu, Li, et al., 2026). In this study, “multimodal AI teaching assistant” is defined as "education and teaching assistant system under the background of generative artificial intelligence"□ with the help of cross-modal and multi-modal fusion generative big language model, as well as other artificial intelligence education agent software, through real-time processing of multi-source heterogeneous data in teaching, personalized teaching support, intelligent evaluation and feedback are provided for teachers and students to complete the iterative optimization of teaching process. Based on this, this study takes the deep integration of "multimodal teaching" and "AI teaching assistant" as the background, focuses on the core research framework of "design-implementation-feedback-optimization" closed-loop, and focuses on how AI teaching assistants can improve the efficiency of lesson preparation, classroom effect, refinement of evaluation feedback and optimization of teaching closed-loop.

3.MULTIMODAL AI TEACHING ASSISTANTS FOR INSTRUCTIONAL DESIGN

Multimodal AI assistant teaching design refers to the integration of language, images, sound and other symbols with a variety of AI tools or platforms to maximize the expression of meaning and guide students to cooperate and communicate based on multiple expressions.

Traditional college English lesson preparation mainly has the following difficulties : 1) There is a significant difference between students ' English foundation and learning motivation, and the unified teaching design is difficult to adapt to the hierarchical learning situation ; 2) The course needs to take into account the cultivation of language examination ability and the development of humanistic critical thinking literacy, and the teaching orientation is not easy to balance ; 3) Multimodal teaching resources are complicated, material screening and curriculum integration are difficult, and multimedia forms are more likely to appear than teaching effectiveness; 4) Large class teaching restricts the implementation of interactive tasks, and classroom activities are easy to flow on the surface.

In view of the above difficulties, the multi-modal AI assistant teaching design model can provide flexible and targeted response methods. In the lesson preparation stage, teachers can make full use of various AI tools to collect multimodal resources. Multimodal resources are not simple teaching decoration, but an important

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

carrier for students ' cognitive processing. By combining images, audio, short videos and other materials with language knowledge points, through multi-modal situation construction, the memory and understanding of language points can be deepened in multi-sensory participation. It can help students jump out of the limitations of pure text learning, complete meaning generation in situational awareness, and lay a good foundation for optimizing teaching effect (Liu, S., 2025).

After collecting multimodal teaching resources, teachers can use AI assistants to design hierarchical teaching plans and implement hierarchical teaching design for the teaching links in the teaching process in combination with AI applications. Teachers can make full use of AI tools to create authentic contexts and design teaching activities with both interactive teaching mode and situational teaching mode (Jia, 2022). For example, when designing topic discussions to introduce course topics, a considerable number of students may have learning anxiety due to limited oral foundation. With the help of AI models such as DeepSeek and beanbags, this activity can be transformed into a teaching design for different students. Take the lead-in activity of Lesson 1 in New Horizon College English Reading and Writing Course 1 as an example: “Work in groups and discuss what ' a brighter future ' means to you.”

The original activity form is relatively simple, not suitable for hierarchical learning, has no relevant context, and has no interactive design. With the help of the AI dialogue model, the prompt words can be continuously optimized, and the original design can be transformed into teaching steps with situational and interactive details.

For example, after applying the DeepSeek prompt (“you are a college English teaching consultant, please redesign this college English classroom activity based on different oral foundations (high, medium and low)”), the following hierarchical design can be obtained :

Table 1: AI-generated Multimodal Teaching Instructions

Foundational Level	Intermediate Level	Advanced Level
The teacher uses an AI image-generation tool to create five distinctive images of future scenarios: graduates tossing their caps, professionals in suits shaking hands, three generations sharing a meal, volunteers teaching in rural communities, and a rocket lifting off. The teacher plays different styles of music and asks students to identify the image prompt suggested by each piece. 3. Using a sentence frame, students say: “To me, a brighter future is about [word corresponding to the image].”	1. Students select from the teacher-provided images the factor they consider most important. 2. The teacher opens an online interactive platform, and five options are projected on the main screen. The class votes anonymously on the question: “Which factor is most important for a ‘brighter future’?” A bar chart or word cloud is generated instantly to show to all. 3. The teacher encourages students to explain their choices using the scaffold: “To me, a brighter future means having ___. Because ___.”	1. The teacher poses challenging questions to encourage dialectical thinking and uses AI to generate concrete images of conflict (e.g., career success accompanied by family absence). 2. Using language scaffolds, students state their positions and compare them with those of others: "From my perspective, ... ranks first because ...” and “I see your point, but compared with X, I think Y is more fundamental.”

From the above teaching design, it is not difficult to find that the application of AI assistants in lesson preparation can not only significantly reduce the time spent by teachers in lesson preparation, but also easily realize the teaching design of stratification + situation + interaction for the existing pain points of college En-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

glish lesson preparation.

College English teaching for non-English majors also faces another practical problem : how to realize the organic integration of CET-4 examination-oriented training and the cultivation of humanistic critical literacy. Taking the vocabulary lesson preparation as an example, teachers can use the AI tool to orient multimodal design and connect the four-level test points at the same time, taking into account the humanistic thinking-oriented example sentences and supporting exercises, so as to take into account the examination needs and literacy development goals. For example, the word "triumph" can use AI tools to set up four-level questions such as word meaning discrimination and sentence translation around the word, and at the same time build a context with themes such as youth growth and home-country struggle to help students understand the triumph. The deep connotation of achieving a major victory after a difficult struggle inspires students to carry out humanistic thinking on the basis of consolidating their examination-related language skills.

In traditional lesson preparation, in order to obtain better teaching effects, it is necessary to use various multimodal teaching resources. However, how to configure it to achieve the optimal effect can only be determined according to the different teaching styles of teachers. Although Zhang Delu put forward the best effect principle of multimodality (effective principle, adaptation principle and economic principle) in 2010, it is not detailed and systematic enough (Zhang, D., 2010). In the context of the current widespread use of smartphones and social media, students' preference for non-verbal modalities has become an important practical reference to be considered in lesson preparation. Non-verbal modes can stimulate students ' interest in learning and sense of participation. Teachers can use non-verbal modes as a starting point to encourage students to achieve "the fit of their own conceptual meaning and language form" (Liu, S., 2023). Therefore, combining students ' own meaning construction and making full use of non-verbal modes to carry out teaching design is the core practice path of current multimodal teaching. For example, still taking the word triumph as an example, teachers can use AI tools to generate a video in advance in the preparation of lessons (such as the climber 's efforts to climb to the top, and let the climber use triumph to describe the difficulty of climbing to the top and the thinking after success. In classroom teaching, teachers can let students watch videos, and then encourage students to recall scenes in their own experiences that can use trumph, use AI tools to generate pictures or short videos, and then create sentences with the help of AI tools to describe them. This kind of teaching design can not only enable students to fully experience the context of word use, but also better complete the test-oriented ability training requirements and humanistic thinking.

In the face of the dilemma that interactive tasks are difficult to land and classroom activities are superficial under large class teaching, teachers can rely on AI assistants to set up multi-modal hierarchical collaborative tasks, complete with clear evaluation brackets, combine online asynchronous discussion with classroom short-term display, and take group output as the starting point to promote students ' deep participation. For example, in the design of unit vocabulary teaching, teachers can design different tasks at the basic level and the promotion level, and promote different basic students within the group to complete active participation and thinking at different levels. The students at the basic level study the multimodal materials such as news short videos and pictures provided by AI, sort out the meaning of words, and sort out the four-level examples. Students at the promotion level combine the humanistic connotation or ideological and political theme behind the material mining, and use AI to complete short multimodal group works (graphic short copy / dubbing

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

small fragments).

4.MULTIMODAL AI TEACHING ASSISTANTS FOR CLASSROOM EFFECTIVENESS, ASSESSMENT, AND FEEDBACK

Multimodal 'AI assistants' can help classroom teaching effectiveness and evaluation feedback by constructing a ternary symbiosis model of ' teacher-multimodal AI assistant-student '. That is, through various AI applications and generative big models, big data analysis and teacher-student collaborative interaction can effectively solve the contradiction between scale and personalization in traditional teaching (Zhang, L., & Ouyang, 2026).

First of all, the multimodal AI teaching assistant can be used as a data hub to connect the whole process of listening and speaking, reading, writing and translation. In listening and speaking teaching, AI relies on multimodal situations to carry out pre-assessment, collect learning behavior and answer data, assist teachers in diagnosing academic conditions, generate multimodal hierarchical listening and speaking tasks, and summarize common problems in the classroom to support targeted classroom evaluation. In the teaching of reading, writing and translation, AI can generate multi-modal hierarchical reading texts, build multi-modal scaffolding for writing, collect asynchronous learning data, automatically identify basic language errors, and help teachers locate key and difficult points and implement differentiated teaching.

Secondly, evaluation feedback can adopt a human-machine coordination mechanism : AI is responsible for objective evaluation and learning process traces, and teachers focus on higher-level literacy, such as argumentation logic and humanistic speculation, to carry out comments. Under this symbiotic framework, teachers assume the role of multimodal teaching design and high-level empowerment. AI assistants are responsible for providing learning diagnosis, resource push and process evaluation in a multimodal way. Students rely on various multimodal learning data to independently plan their learning paths, which not only alleviates the pressure of teaching evaluation in large classes, but also realizes the unity of objectivity and humanity in intelligent evaluation.

In the construction of hierarchical evaluation feedback language, “Multimodal AI Assistant” assists teachers to prefabricate instant feedback speech and multimodal comment strategies for each module of listening, speaking, reading, writing and translation, and teachers rely on the preset framework to implement personalized classroom comments ; students combine hierarchical feedback to independently diagnose deficiencies, adjust learning paths, and form an evaluation closed-loop. For example, in the oral class, teachers can use prompts to call AI to generate a summary table, integrate weak problems, short and quick comments, encouraging error correction and multi-modal reinforcement strategies, serve classroom layered and instant comments, and promote the symbiosis between teachers and students. Taking oral classroom feedback as an example, give Doubao prompt word : “Make a summary table of the instant feedback language of classroom teachers in the five modules of listening, speaking, reading, writing and translation, including weak problems + short quick comments + encouraging corrections.” The evaluation table (taking listening and speaking as an example) is generated as follows:

Table 2: AI-generated Multimodal tired feedback

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Skill	Areas for Development	Immediate Classroom Feedback	Multimodal Reinforcement Strategy
Listening	1. Understands the main idea but misses key details, numbers, or logical markers 2. Cannot keep pace and becomes anxious or disengaged 3. Stops processing when encountering unfamiliar words 4. Recognizes only the original wording, not paraphrases	1. Great main idea! Just catch more key details next time. 2. Good try! Stay calm and follow the rhythm. 3. Nice word catching! Learn to guess instead of stopping. 4. Nice observation! Try to notice paraphrasing next time.	1. Projected highlighting 2. Adjustable-speed audio 3. Illustrated vocabulary cards 4. Side-by-side text comparison
Speaking	1. Willing to speak but frequently hesitates or pauses for a long time 2. Gives brief, underdeveloped responses without elaboration or reasons 3. Makes minor grammatical errors (e.g., tense or number) that do not impede understanding 4. Pronounces words accurately but uses flat, read-aloud intonation 5. Speaks fluently but repeats a limited range of sentence patterns	1. Brave speaking! Let's make it smoother next time. 2. Clear opinion! Add more details to enrich your speech. 3. Good expression! Fix tiny grammar slips next time. 4. Nice pronunciation! Try more vivid intonation. 5. Very fluent! Let's diversify your sentences.	1. Pop-up sentence scaffolds 2. Mind maps 3. Color-coded annotations 4. Audio models 5. Sentence-transformation comparison chart

Multimodal AI teaching assistants can easily refine the implementation strategies of classroom teaching activities and provide strong support for teachers ' flexible classroom teaching and accurate evaluation feedback. With the rapid development of AI technology, it provides more experimental space for multimodal real-time conversion in classroom teaching activities of listening, speaking, reading, writing and translation. For example, in classroom teaching, AI tools and platforms can be used to interact in various forms (barrage, word cloud and online voting, etc.).), meaningfully generate multimodal teaching activities, activate classroom atmosphere, and help students understand learning content in multiple dimensions (Ma, 2024).

5.CONSTRUCTING A MULTIMODAL AI-ASSISTED TEACHING FRAMEWORK

Constructing a multimodal AI-assisted teaching framework is an important means of optimizing the instructional cycle. Supported by generative AI, a structured system that integrates multiple information modes and uses human–AI collaboration to generate multimodal resources intelligently, implement generative learning activities, and complete a closed instructional loop can enable personalized smart teaching at scale. Specifically, the framework comprises a complete cycle: teacher input of instructional requirements → batch generation of multimodal resources by an AI teaching assistant → generative instructional implementation before, during, and after class → collection and analysis of learning data → iterative optimization of framework rules. Such a generative multimodal framework is adaptive, reusable, and scalable.

Amid rapid advances in AI, "multimodal teaching-resource generation frameworks can use AI-based conversational programming to generate resources on demand with a single action and thereby create instruction-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

al resources at scale" (Ou et al., 2024). Through general-purpose large-language-model interfaces, teachers can use natural-language dialogue to drive a web-based resource-generation engine. Without specialized programming skills, they need only enter a topic, difficulty level, and learning objectives to produce, in batches, multimodal texts, listening scripts, tiered tasks, and courseware suited to College English classrooms. This can overcome the long production cycles, limited variety, and inadequate personalization of conventional resource development.

The construction of a multimodal teaching framework necessarily depends on the participation of learners themselves. Continually advancing multimodal intelligent-interaction technologies now provide robust support for a learning cycle of "input–analysis–feedback–revision" (Huo, 2025). Teachers should remain informed about the latest multimodal intelligent-interaction technologies, adapt to changing roles, and guide students in constructing adaptive and personalized learning pathways.

The deep integration of AI into multimodal teaching requires researchers to "bring macro-level concerns into the exploration of micro-level features" and to "construct intelligent analytical methods for multimodally integrated instructional structures so as to automate and systematize the identification and quantitative calculation of their spatiotemporal characteristics" (Zhang J., Yu, et al., 2026). Research on multimodal teaching frameworks must therefore investigate the details of the "design–implementation–feedback–optimization" loop while also strengthening the AI literacy of teachers and students. Through their joint efforts, multimodal teaching can move from the conventional accumulation of materials toward an intelligent, precise, personalized, and deeply empowering model.

6.CONCLUSION

Research on multimodal AI-assisted teaching in College English has progressed from initial discussions of principles to innovative models addressing listening, speaking, reading, writing, and translation. It has thus evolved from macro-level theoretical explanation to micro-level classroom implementation and from uniform instruction to differentiated and precise development. The multi-modal teaching mode under the deep integration of "AI Teaching Assistants" is based on the improvement of AI literacy of teachers and students. With the advantages of different teaching modes and various AI tools carrying out multi-dimensional and detailed research on the "design-implementation-feedback-optimization" teaching closed-loop, a new multi-modal teaching paradigm is hoped to be established. This is a step from empirical and homogeneous teaching mode to digital, differentiated and precise intelligent education mode.

Funding

This article is an outcome of the 2025 "Guangdong Provincial Undergraduate Universities Special Project on Foreign Language Teaching Reform Research and Practice", entitled "An Empirical Study of the Effectiveness of College English Supported by Multimodal AI Teaching Assistants" (Project No. 25GWYB17).

本文系 2025 年广东省本科高校"外语教学改革研究与实践"专项课题"多模态 AI 助教赋能大学英语效能实证研究"（项目编号：25GWYB17）的研究成果。

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

References

[1]Alitunur. (2026). Multimodal AI technology facilitates interactive design of college English classrooms. Teaching of Forestry Region, (6), 117–120.

[2]Huo, N. (2025). Applications and effects of multimodal intelligent-interaction technologies in college English teaching. English Square, (24), 83–86. https://doi.org/10.16723/j.cnki.yygc.2025.24.027

[3]Jia, H. (2022). Multimodal information-based college English teaching in the age of artificial intelligence. Journal of Inner Mongolia University of Finance and Economics, 20(1), 30–33. https://doi.org/10.13895/j.cnki.jimufe.2022.01.030

[4]Kress, G., & van Leeuwen, T. (2001). Multimodal discourse: The modes and media of contemporary communication. Arnold.

[5]Kress, G., & van Leeuwen, T. (2006). Reading images: The grammar of visual design (2nd ed.). Routledge.

[6]Liu, J. (2023). A case study on innovative pedagogy of college English—from a multimodal perspective. Overseas English Testing: Pedagogy and Research, 5(1), 14–22. https://doi.org/10.12677/OETPR.2023.51002

[7]Liu, S. (2025). Optimization strategies for online college English teaching in multimodal interactive environments. Journal of Jiamusi Vocational Institute, 41(3), 139–141.

[8]Lu, H., Li, X., Zheng, L., Gong, T., & Hu, L. (2026). Practices and prospects of generative AI teaching assistants. Software Guide, 1-7.

[9]Ma, X. (2024). Research on innovative multimodal informationization teaching of college English in the era of Artificial Intelligence. Journal of Hubei Open Vocational College, 37(11), 9–10.

[10]Ou, Z., Liu, Y., Qin, K., Wang, Y., & Li, X. (2024). Generation and evaluation of artificial intelligence multimodal teaching resources—based on the application of AIGC in international Chinese language education. Modern Educational Technology, 34(9), 37–47.

[11]Wang, Y. (2024). Study on AI Zhihui multimodal college English writing workshop. Modern English, (23), 67–69.

[12]Wang, Z., & Chen, Y. (2026). Evolving research hotspots and future trends of multimodal teaching at home and abroad: A CiteSpace knowledge-graph analysis. Educational Information Technology, (Z2), 132–136+89.

[13]Yang, P. (2019, October 10). Using big data and artificial intelligence to guide research on a multimodal assessment system for college English writing proficiency. In Proceedings of the 16th Shenyang Annual Scientific Conference (Economics, Management, and Social Sciences). The 16th Shenyang Annual Scientific Conference, Shenyang, Liaoning, China.

[14]Zhang, D. (2010). Preliminary investigation into the concept of design and the selection of modalities in multimodal foreign language teaching. Foreign Languages in China, 7(3), 48–53. https://doi.org/10.13564/j.cnki.issn.1672-9382.2010.03.011

[15]Zhang, J., Yu, X., Chen, Z., & Yang, Y. (2026). Intelligent analysis method for multimodal integrated teaching structure. Modern Educational Technology, 36(6), 64–73.

[16]Zhang, L., & Ouyang, J. (2026). Innovation in university teaching models from the perspective of symbiosis theory: The case of English teaching. Higher Education Exploration, (4), 96–101.

[17]Zhou, C. (2025). A Study on multimodal college English reading teaching model enabled by artificial in-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

telligence technology. Modern English, (1), 7–9.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)