US-China Education Review A, July 2025, Vol. 15, No. 7, 526-532
doi: 10.17265/2161-623X/2025.07.006

On the Integration of GAI and Multi-modality in English Teaching: A Study on New Language Teaching Model Construction

LI Kunmei, HU Kun
Nanfang College Guangzhou, Guangzhou, China

This article systematically integrates the powerful generative features of Generative Artificial Intelligence (GAI) with the synergistic features of multi-modal technology and explores a new pedagogical approach to effective language teaching under the context of a lack of active engagement and motivation, and limited accessibility and dynamism of educational materials in traditional English courses in China's advanced education. This study proposes a novel language teaching model (a three-tier structure) based on a critical review of the usage of GAI and multi-modality in educational environments. This new language teaching model combines GAI with multi-modal technology and centers around the G-M4 cycle (Generation-Input-Interaction-Output-Monitor/Feedback). This model means empowered generative capabilities, a more dynamic and interactive learning environment, multi-modal and creative output, and effective evaluation and prompt feedback. Furthermore, critical aspects that require attention, such as data privacy and ethical responsibilities, are also illustrated.

Keywords: Generative Artificial Intelligence, multi-modal technology, English teaching, GAI integration

Introduction

Multi-modal teaching theory fundamentally reshapes language pedagogy by coordinating visual, auditory, and tactile channels to scaffold knowledge construction (Kress & van Leeuwen, 2020). Currently, with the advancement of new technologies and educational tools, Generative Artificial Intelligence (GAI) is increasingly driving educators to rethink the overall language education ecosystem. Undoubtedly, it has attracted widespread attention from the academic circle due to its powerful language processing as well as content creation capabilities. However, current research on GAI and language teaching still disproportionately focuses on isolated technical applications, such as GAI-assisted English writing and GAI-assisted speaking drill, with research in systematic integration of GAI's powerful generative features and the advantages of multi-modal praxis still lacking. Therefore, this study aims to bridge this gap and explore the possibilities of deep integration between the two. This paper proposes an innovative teaching model that integrates GAI with multi-modality, a new language teaching model based on a three-tier structure with the concept of "Generation-Input-Interaction-Production-

Acknowledgments: This work is supported by the 2025 Collaborative Education Project of Industry-Academia Cooperation (Grant Number: 2506184205): "Exploration of Teaching Reform Path of Advanced English Course Under the Perspective of AI Empowerment".

LI Kunmei, Lecturer, School of Foreign Languages, Nanfang College Guangzhou, Guangzhou, China.
HU Kun, Lecturer, School of Foreign Languages, Nanfang College Guangzhou, Guangzhou, China.

Monitor/Feedback" (G-M4) in the middle tier being the structural core. This model is supposed to provide a new path for enhancing language learners' comprehensive language application abilities, cross-cultural communication skills, and learning motivation, and provide pedagogical justification for empowering language education with informational as well as multi-modal literacy.

Literature Review

Generative Artificial Intelligence (GAI) in Education

With the Age of Information gradually transforming into the Age of Intelligence, GAI is also on a penetration process not only into people's daily life, but also into other significant areas, including education. Though certain governments have made their digitalization policies regarding GAI integration in educational practices, teachers and educators have shown their awareness and perception of the usage of GAI as tools. GAI has shown pedagogical significance by generating dynamic teaching materials and offering customized learning adaptation, such as generating scenario text, providing sample writings, collecting reading materials and targeted exercises that match the learner's language level (Kasneci et al., 2023). Scholars in China have found that AI-based personalized reading materials can significantly affect language learners' reading fluency and vocabulary acquisition levels (Huang, 2023; Du & Gao, 2022). Furthermore, GAI can act as a virtual language partner, providing learners with real-time grammar and vocabulary correction suggestions (Kohnke, Moorhouse, & Zou, 2023).

Multi-modality in Language Teaching and Practice

Multi-modality emphasizes practical applications. Language, as a social symbol, together with other modalities such as images, sounds, movements, spatial arrangements, etc., jointly constructs meaning because people convey meanings not just through words or sounds. The multi-modal educational practice emphasizes the collaborative meaning-making function of language symbols and non-language symbols (such as images, sounds, movements, space) in social and cultural contexts (Jewitt, 2009; Zhang & Li, 2023). Multi-modal teaching, by utilizing various symbolic modes such as texts, images, audios, and videos, aims to integrate multiple sensory channels and create a more authentic language environment, thus significantly enhancing students' engagement and depth of understanding (Serafini, 2014). Currently, the advancement of digital technology has greatly expanded multi-modal resources, making such multimodal tools like interactive whiteboards, VR/AR, digital storytelling, etc., with easier accessibility. In conclusion, multi-modal input can effectively reduce cognitive load and enhance the comprehensibility of language input (Mayer, 2020), while multi-modal output promotes the deep internalization of language meaning and creative expression (Zhang & Li, 2023).

Research Gap in Integration of GAI and Multi-modality in Language Teaching

Present GAI research predominantly examines GAI's textual output capacities and the use of multi-modal resources in language pedagogy. However, systematic investigation of GAI with multi-modality to advance instructional design remains underdeveloped (Wang, Liu, & Su, 2025). Responding to this gap, this paper proposes a new English teaching model that combines GAI and multi-modality, presenting a three-tier structure with G-M4 at its core. This paper will address: How can teachers utilize GAI's generation ability to diversify multi-modal teaching resources while successfully maintaining pedagogical intentionality? How can educators utilize GAI-based multi-modal inputs to gain authentic learner interactions and achieve meaning-laden outputs and what are critical challenges that require further attention in the multi-modal evaluation process?

Three-Tier Teaching Model: Construction and Innovation

The model presented in the figure below is based on the 2025 Official Guidelines for Safe Use of Generative Artificial Intelligence in Education issued by China's Ministry of Education, as well as guidelines for students' use of generative artificial intelligence from various universities in China, including East China Normal University and Shanghai Tech University. This learner-centered, task-driven, and GAI-assisted model is depicted in the figure below.

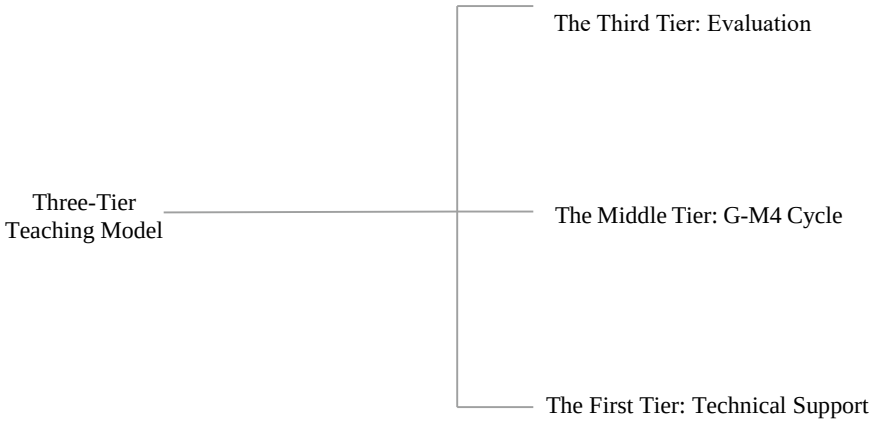

Figure 1. Diagram of three-tier model.

In the above structural model, the first tier focuses on technical support, with GAI taking the role of technical tool and pedagogical mediator. GAI can construct dynamic scripts, inquiry prompts, evaluative feedback, etc., forming a multi-modal resource library that includes images, audio, video, interactive simulations, VR/AR scenes, and so on. Additionally, it can incorporate a digital analysis platform for capturing learning behavioral data, conducting educational data gathering, and advancing educational research praxis (Mills, 2016). Additionally, it supports personalized recommendations and procedural evaluation.

The mediating tier in the diagram is the model's pedagogical nucleus, with the G-M4 instructional cycle spanning five interdependent phases: G (Generate), M1 (Multi-modal Input), M2 (Multi-modal Interaction), M3 (Multi-modal Output), and M4 (Monitor & Feedback). Please refer to the diagram below:

G (Generate) is a crucial stage of multi-modal contextualized content generation in teaching scenarios. Based on learning objectives, learners' language levels, and interests, GAI can assist language teachers by generating multi-modal stimuli related to the teaching theme and context, for example, if the teaching theme is "technology ethics", GAI can generate a script which may include contents like ethical dilemmas, data privacy breaches, moral implications that are closely related to the application of artificial intelligence. This generated script presents different viewpoints that can be structured in a way that is effective in cultivating students' critical thinking.

M1 (Multi-modal Input) refers multi-modal content input. Teachers or educators can use high-quality multi-modal resources such as audiovisual tools and VR (Zhang & Li, 2023) to empower their teaching instructions with no need of programming skills. These tools are aligned with educational theories like constructivism which holds that learners learn by constructing their understanding and knowledge through real experience and reflection. By combining different types of multimodal tools, teachers can produce creative work and then direct

learners to the immersive input, activating multiple sensory channels to achieve immersive perception and understanding. For example, teachers can produce a multimedia presentation of a scenic spot by making use of GAI and multimodality, and learners can not only see the scenery, but also learn about the daily life of the people.

M2 (Multi-modal Interaction) involves active engagement between learners and GAI, which means students are not just passive knowledge recipients but also active participants in multi-modal dialogues. For instance, learners can describe images or scenes they see. They can describe the landscape, the colors, and their feelings about it. GAI can generate corresponding text based on their speech output and provide an audio feedback by reading out aloud the corrected text form with proper pronunciations. Additionally, learners can execute procedural tasks in a simulated environment just by following audio instructions, as GAI can assess the accuracy of their actions and guide them towards the correct procedures. For example, in writing a business email, GAI can offer a step-by-step guide to students and offer help in the review—and—edit phase if there are any inappropriateness.

M3 (Multi-modal Output) focuses on the construction and production of meaning through the application of linguistic and non-linguistic modes. For example, with the assistance of GAI, learners may engage in tasks that require them to create narrative performances based on AI-generated written scripts, participate in debates using computationally sourced video or audio evidence, and collaborate on theatrical productions which may pose challenges to creative boundaries. Learners at this stage are positioned to be designers of multi-modal meaning as they are the actual center of the G-M4 cycle.

M4 (Monitor & Feedback) refers to intelligent supervision and dynamic feedback. GAI is an unsupervised or partially supervised system that seeks to mimic human behavior and mental processes by making use of artificial digital content such as images, texts, videos, etc. Based on GAI's analysis of learners' data in the G-M1-M2-M3 stages, M4 can provide automated analysis suggestions that span language accuracy, content coherence, and multi-modal coordination. At the same time, teachers, after combining AI analysis of students' data and multi-modal works, provide deep feedback in dimensions such as critical thinking, cultural awareness, and collaboration, forming a loop of "AI fast feedback + teacher deep feedback".

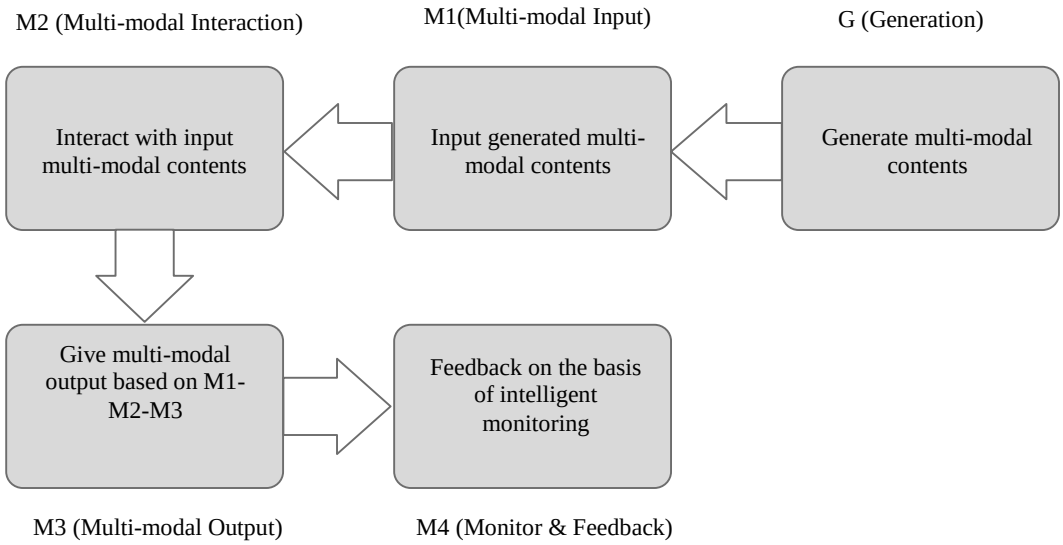

Figure 2. Diagram of the G-M4 cycle.

The ultimate tier is also called the evaluation layer. A multi-dimensional assessment is realized mainly on the ground of proper integration of GAI and multi-modal features, especially by combining process analytics such as interaction patterns, language complexity, error trajectories, etc., with qualitative evaluations made by teachers on dimensions like creativity, cultural salience, and collaborative depth of multi-modal outputs. Of course, in addition to dialectical assessment, this layer of the structure is also involved in addressing ethical and safety considerations brought about by generative artificial intelligence. Transcending technical metrics, it brings into light common concerns such as academic integrity, data privacy, and security.

As an intelligent generator of multi-modal contents, GAI can transcend the traditional static limitations of textbooks by producing dynamic and personalized pedagogical contents, e.g., it can generate up-to-date news for learners to read and discuss, and make responsive adaptation possible in language education, for example, if the language setting is business, GAI can produce contents that are related to business. Building on the multi-modal cycle of Generation-Input-Interaction-Output-Monitor/Feedback, G-M4 deeply integrates GAI's content generation capabilities into the whole process of multi-modal input, interaction, and output, creating a highly authentic and interactive language learning environment.

The proposed teaching model also has the potential to achieve innovation in intelligent assessment. Besides providing real-time, accurate, and prompt feedback at the linguistic level, GAI can help free teachers focus more on cultivation of critical thinking, cultural understanding, and so on, thus reshaping the value of education (Wang et al., 2025). Meanwhile, this model will also integrate and incorporate the cultivation of ethics and critical thinking. Teachers can internalize technical ethics and critical thinking as elements of the teaching mode through task design. For example, teachers can cultivate students' critical thinking and artificial intelligence literacy by having them analyze AI-generated content and discuss AI ethical issues.

Potential Application Scenarios

Scenario 1: Argumentative writing topic: Pros and cons of working at home:

First, in the G (Generate) phase, the GAI is used to generate a multi-modal viewpoint library on the controversial topic of the pros and cons of working at home. This library may include text summaries, info-graphics, expert interview excerpts, and so on. Next, in the M1 input and M2 interaction phase, students/learners analyze the generated materials. Students can engage in debate-style interactions with the GAI to obtain opposing viewpoints, thereby deepening their understanding and knowledge. Based on the deepened understanding and knowledge in M2, students then move on to the M3 output/production phase, where they integrate selected visual and textual evidence. They can even use assistive tools to create a draft of their argumentative essay. The language and logical analysis of the draft are largely provided by the GAI, while students have to conduct crucial data analysis and generate their personalized viewpoints relating to the pros and cons of working at home. In the M4 monitor and feedback stage, GAI provides feedback on the language and logic of the draft, while teachers offer guidance on high-level thinking and in-depth content evaluation based on the students' draft.

Scenario 2: An interview: A Chinese student attending an interview at a UK university:

First, in the generative phase, GAI tools can be applied to mimic the interview scenario and generate AI interviewers. The AI interviewers may be with different personalities, for example, one may be stern-looking, while another may be friendly. In the M1 input and M2 interaction phase, learners engage in immersive interviews with AI simulated interviewers in the VR environment. GAI may generate a wide range of interview questions, and the learners have to learn to give timely response. In the M3 output phase, learners complete responses to

their interview questions. They have to think clearly and organize their thoughts well and give accurate expressions. In the M4 phase, GAI can provide real-time assessment of learners' linguistic appropriateness, e.g., grammar, pronunciation, etc., and cultural adaptability, e.g., giving attention to potential cultural conflicts in the target culture if needed, generate dialogue analysis reports, e.g., strength and weakness of the learner's performance, and offer improvement suggestions and resolution strategies.

Of course, the new teaching mode proposed by the author will certainly face challenges in dimensions of technical costs, educator's GAI literacy, risks of over-reliance, unstable quality of output contents, and potential ethical issues, etc. (Huang, Hew, & Fryer, 2023). However, with the unstoppable advent of technology, especially with the distribution and promotion of open source large language models (LLMs) and low-cost GAI tools as well as multi-modal tools, the technical costs are tending to decrease in the long run. As for GAI literacy, many educational organizations as well as workers are already taking steps to enhance teachers' GAI literacy, for example in China, multiple training courses are being conducted to help teachers learn to use GAI tools and multimodal teaching materials. Generative artificial intelligence will continue to be a transformative force in education, but the principle of "AI as assistance, human as center" in educational design will not change, and the essence of technology serving education will hold.

Conclusion

This new three-tire language teaching mode is innovative in that it integrates GAI's powerful content output ability with multi-modal technology and provides a comprehensive G-M4 approach to language education. The cycle of Generation-Input-Interaction-Output-Monitor/Feedback is significant in creating a personalized and interactive language learning environment. This model not only provides an effective path to improve language skills but also demonstrates unique value in cultivating learners' critical thinking, cross-cultural awareness, and digital literacy. In the future, the integration of GAI and multi-modality will be promising and new possibilities of technology-empowered language education will surely open up. Confronted with technological upgrading and educational evolving, we should be, on the one hand, active explorer to advanced technology and enhance overall GAI literacy. However, on the other hand, we should also cautiously apply cutting-edge technology, guard against excessive reliance on it, and be extremely cautious about the potential ethical risks brought about.

References

Du, Y., & Gao, H. (2022). Determinants affecting teachers' adoption of AI-based applications in EFL context: An analysis of analytic hierarchy process. *Education and Information Technologies, 27*(7), 9357-9384. Retrieved from https://doi.org/10.1007/s10639-022-11001-y

Huang, F. (2023). Examining foreign language teachers' information literacy: Do digital nativity, technology training, and fatigue matter? *The Asia-Pacific Education Researcher, 33*, 901-912. Retrieved from https://doi.org/10.1007/s40299-023-00797-z

Huang, W., Hew, K. F., & Fryer, L. K. (2023). Generative AI and the future of education: Ragnarök or reformation? *Educational Technology Research and Development, 71*(1), 1-14. Retrieved from https://doi.org/10.1007/s11423-023-10231-2

Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. London: Routledge.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F. … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. Retrieved from https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(3), 537-550. Retrieved from https://doi.org/10.1177/00336882231162873

Kress, G., & van Leeuwen, T. (2020). *Multimodal discourse: The modes and media of contemporary communication* (2nd ed.). London: Bloomsbury Academic.

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/us/univerisitypress/subjects/psychology/educational-psychology/multimedia-learning-3rd-edition

Mills, K. A. (2016). *Literacy theories for the digital age: Social, critical, multimodal, spatial, material and sensory lenses*. Bristol: Multilingual Matters. Retrieved from https://www.multilingual-matters.com/page/detail/?k=9781783094615

Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. New York: Teachers College Press. Retrieved from https://www.tcpress.com/reading-the-visual-9780807754719

Wang, Z. J., Liu, X. J., & Su, C. Y. (2025). How humans differ from artificial intelligence: A design thinking-based approach to cultivating empathy in an AI-driven society. *Journal of Distance Education*, *45*(6), 20-34. Retrieved from https://doi.org/10.13541/j.cnki.chinade.2025.06.005

Zhang, Z., & Li, J. (2023). Multimodal input in second language learning: A meta-analysis. *Language Learning & Technology*, *27*(1), 1-24. Retrieved from https://doi.org/10125/73518