International Journal of Social Science, Education and Humanities
ISSN Print: 3104-4239
ISSN Online: 3104-4247
CD ISSN: 3104-4255
https://doi.org/10.63313/IJSSEH.9034

Human-Machine Collaboration: Boundaries and Practical Paths of Generative AI in Spoken English Teaching

Xiaolan He
School of Foreign Languages, China West Normal University, China
Email: 1485236302@qq.com

How to cite this paper: He, X. L. (2026). Human-machine collaboration: Boundaries and practical paths of generative AI in spoken English teaching. *International Journal of Social Science, Education and Humanities*, *2*(2), 34–43. ISSN Print: 3104-4239; ISSN Online: 3104-4247.
https://doi.org/10.63313/IJSSEH.9034
Published: 2026-05-14

Abstract

Grounded in the philosophy of technology and humanistic pedagogy, this study focuses on the application boundaries of generative AI in oral English teaching, aiming to clarify the division-of-labor logic in human-machine collaboration. Based on Stiegler's "technological prosthetics" and "pharmacology" models, as well as humanistic teacher-centered theory, the study reveals that the auxiliary value of generative AI lies in mechanical, standardized, and low-culture-loaded tasks. In contrast, the uniqueness of human teachers is rooted in domains beyond AI's reach, such as transmitting emotional resonance, deeply interpreting cultural contexts, and guiding dynamic thinking. A three-dimensional boundary criterion and a four-quadrant model are proposed to define the practical scope of collaboration, where technical proficiency determines the feasibility of AI intervention, while the irreplaceability of teachers defines the necessity of their roles. Additionally, a dual-path approach is proposed: developing teachers' critical AI and cultural pedagogies, and designing tools with cultural diversity to mitigate bias. The study concludes that the the essential boundary is ethical, safeguarding "human subjectivity and cultural nature," thus offering a framework for responsible integration.

Keywords

Generative AI; Oral English Teaching; Application Boundaries; Human-Machine Collaboration

1. Introduction

With the rapid advancement of generative AI technology, its applications in the field of education are becoming increasingly widespread. Particularly in English speaking instruction, various intelligent tools are profoundly transforming traditional "listening-speaking" interactive scenarios (Feng,2024). Generative AI is gradually permeating every aspect of speaking instruction thanks to its technical advantages

in natural language processing and the generation of context-aware dialogue. However, alongside this technological empowerment, a deep-seated contradiction regarding the boundaries of human-machine roles is becoming increasingly apparent: a tension has emerged between the highly efficient support provided by generative AI and the irreplaceable uniqueness of teachers, resulting in functional overlap and conflict. While AI demonstrates superior efficiency in standardized language tasks, the core of spoken English instruction lies in dimensions inseparable from human agency—emotional resonance, cultural interpretation, and dynamic thinking—areas where teachers remain irreplaceable. This contradiction not only raises concerns about whether "technology will replace teachers," but also compels educators to consider: in English speaking instruction, how can we delineate the scope of generative AI's support and the core domain of teachers in a way that fully leverages the tool-based value of technology while upholding the humanistic essence of teaching, thereby avoiding the risk of "dehumanization" caused by the misuse of technology? By constructing a tripartite boundary model, this paper delineates the complementary roles of AI and teachers in spoken English instruction, thus addressing the fundamental need to balance technological efficacy with the irreplaceable human dimensions of learning.

2. A Framework for Analyzing the Human-Machine Boundary in Oral Language Instruction

2.1. The Prosthetic Nature of Technology and the Pharmacological Model

Stiegler's models of "Technological Prosthesis" and "the Pharmacological Model" provide key theoretical perspectives for understanding the role of generative AI in English speaking instruction (Stiegler, 1998; Stiegler, 2013). In particular, the concept of "technological prosthetics" emphasises technology as an extension of the human body and capabilities, expanding a person's range of action much like a "prosthetic limb". In the context of spoken English teaching, this characteristic manifests specifically in generative AI's efficient replacement of "mechanical tasks"—leveraging its technical capability for "memory-based dialogue generation", AI can accurately perform repetitive tasks such as pronunciation repetition drills and simulated conversations on fixed topics. For example, it can use speech recognition technology to repeatedly correct students' pronunciation of phonetic symbols, or generate dialogue scripts for standardised scenarios such as "shopping" for students to practise. This extends the scope of a teacher's capabilities in routine teaching activities and alleviates the burden of repetitive work. The "pharmacological"model, however, reveals the dual nature of technology: whilst it serves as a "remedy" for solving problems, it may also become a "poison" that gives rise to new ones: When English teachers rely excessively on generative AI to perform basic tasks in oral teaching, the "toxicity" of technology may become

apparent—teachers may gradually weaken their core competencies in dynamic interaction. This erosion of teacher agency and creativity in pedagogical moments constitutes the core "toxicity," which aligns with the broader risk of a decline in innovative spirit due to technological intervention (Hou, 2025), thereby threatening the depth of cultural and cognitive engagement in language learning.

2.2. The Humanist Teacher-Centered Approach

The humanist teacher-centered approach posits the teacher as an irreplaceable cultural interpreter, a role that transcends the mere correction of language forms. When students make pragmatic errors, such as literal translations stemming from Chinese-centric thinking, teachers can trace these to their historical and cultural roots, whereas generative AI typically offers only superficial corrections devoid of underlying cultural logic (Gray, 2018). This capacity for in-depth cultural interpretation stems from teachers' embodied understanding of human society and their cross-cultural experience, forming the foundation of their irreplaceability. The fundamental nature of spoken language teaching further reinforces the central role of the teacher. Unlike the standardization dialogue pathways pre-programmed into AI, authentic spoken communication is unpredictable and context-dependent (Richards, 2006). For instance, when students discussing "environmental policies'" suddenly bring up "traditional agricultural wisdom", the teacher can immediately expand this into an interdisciplinary cultural exploration, whereas AI, constrained by its algorithmic framework, rigidly returns to preset pathways. Such interaction is a socio-cognitive practice, reliant on a teacher's immediate feedback—through eye contact, tone, and adaptive questioning—to foster deeper thinking, a process whose emotional warmth and cognitive flexibility exceed AI's formulaic responses (Vygotsky, 1978). As the critique of "banking education"warns, reducing oral language teaching to AI-driven, mechanical output training would undermine the essential function of language as a vehicle for humanistic expression (Freire, 1970).

3. Delimiting the Auxiliary Value of Generative AI

3.1. AI as a Tool for Empowerment

The core value of generative AI in spoken language teaching is primarily reflected in the precise technical support it provides for the training of basic skills, particularly in the areas of speech recognition and real-time error correction. Speech recognition systems based on deep neural networks (e.g., Google's Speech-to-Text API) can achieve near-human-level accuracy in converting speech to text and provide real-time annotation of pronunciation errors (e.g., misplacement of the tongue for vowels or the omission of voicing in consonants). Such technologies can effectively replace traditional repetitive, mechanical drills, freeing teachers from basic error correction so they can focus on cultivating higher-order communicative skills(Huang, Baker, & Reddy, 2022). However, we must remain vigilant regarding

the risks of over-reliance on technology: AI can only identify overt phonetic deviations (e.g., confusion between /s/ and /θ/), but cannot diagnose pragmatic errors caused by cultural factors (e.g.,Chinese students overusing the reduced modal verb "could" due to the principle of modesty). In terms of generating virtual conversational scenarios, large language models (i.e., LLM) demonstrate significant advantages. Models such as GPT-4 or Llama 2 can generate dialogue templates that closely resemble real-world contexts based on prompt engineering (e.g., a scenario involving 'protesting against excess baggage at airport check-in'), and can control linguistic complexity (e.g., vocabulary at B1 or C1 levels). However, existing technologies suffer from limitations in terms of superficial cultural understanding: generated dialogues in Western restaurants may include details on cutlery placement, yet overlook the conflict of labour values underlying tipping culture, whereas such deep-seated cultural logic is precisely the teaching focus for teachers acting as "negotiators of meaning" (Huang, Baker, & Reddy, 2022).

3.2. AI as a Lever for Efficiency

Generative AI significantly optimises the efficiency of teaching resource allocation by automating routine training tasks. In the development of basic skills, such as repetitive pronunciation practice, AI systems utilise Automatic Speech Recognition (i.e., ASR) technology to capture phonetic errors in students' spoken output in real time and generate instant visual feedback. This mechanism of instant feedback increases the frequency of effective practice per unit of time for students, whilst freeing teachers from repetitive monitoring tasks, allowing them to redirect their efforts towards designing higher-order cognitive activities. In expanding personalised practice scenarios, generative AI's "structural text generation" capability plays a pivotal role. Generative AI overcomes the limitations of textbook-bound practice by generating rich, personalised scenarios based on multidimensional student data, including language proficiency, interests, and learning objectives. Taking the well-known online English learning platform Preply as an example, it utilizes large language models to analyze students' past learning trajectories. For instance, for a student interested in travel and at an intermediate English level, it can generate conversation scenarios tailored to their interests and proficiency, such as "planning a backpacking trip across Europe." These scenarios incorporate authentic travel vocabulary and cultural background information. Furthermore, these scenarios are not static; as students make different choices during the dialogue, the AI will intelligently expand dialogue branches based on "structural text generation" logic, simulating the flexible interactions found in real-life communication. This immerses students in authentic travel scenarios, significantly enhancing the relevance and engagement of speaking practice.

3.3. The Non-Negotiable Boundary of Assistance

However much the supportive role of generative AI in spoken English teaching may

be expanded, it must strictly adhere to two inviolable humanistic boundaries; these constitute the core bottom line for ensuring that teaching does not stray from the "essence of education". In terms of emotional interaction, whilst AI can provide standardised feedback (e.g., "Excellent pronunciation!") through pre-set programmes, it remains incapable of capturing and responding to the complex and subtle emotional fluctuations inherent in human dialogue (Rogers, 1969). When students experience "speech anxiety" due to cultural differences, teachers can establish a psychological safe space through body language and contextual reassurance, whereas AI's formulaic responses may exacerbate emotional detachment. The interpretation of cultural contexts further highlights the irreplaceable role of teachers as "negotiators of meaning". Every expression in spoken English is embedded within a specific cultural logic; from the pragmatic choices in everyday greetings to the implicit transmission of values, all require dynamic interpretation within historical and social contexts. For example, when a student asks, "Why do Westerners dislike the number 13?", an AI might list relevant religious references, yet it struggles to analyse, as a teacher would, the interweaving of the Christian concept of original sin and medieval social psychology underlying this taboo; when explaining the differences in the concept of the "dragon" between English and Chinese contexts, an AI might merely contrast the symbolic meanings of "auspicious" and "evil", whereas a teacher can draw upon the philosophy of the unity of heaven and humanity and the Western mythological tradition to guide students in understanding the differences in values underlying these cultural symbols. The ability to provide such deep interpretation, which hinges on the educator's synthesis of historical, social, and psychological knowledge, remains a fundamental challenge for data-driven generative AI. If technology is allowed to dominate cultural interpretation, it risks fostering a superficial understanding of culture and perpetuating stereotypes, thereby contradicting the goal of "cultural responsiveness"—building cross-cultural empathy through understanding, not just memorising symbols.

4. The Irreplaceable: Core Competencies of Human Teachers

4.1. Guidance in Affective and Cognitive Dynamics

The core value of human teachers lies in their ability to create dynamic cognitive-affective spaces, a capacity that generative AI cannot replicate. When students struggle with expression due to cultural differences—for instance, a learner overusing hesitant phrases like "maybe I think…"—a teacher's empathetic non-verbal cues and vocal tone can alleviate anxiety and inspire the courage to communicate more directly. A metaphorical encouragement (Show me your Brazilian passion in English) often sparks breakthrough expressions that blend cultural traits, such as "Samba rhythm beats in my arguments!" This burst of creativity, rooted in the emotional resonance between teacher and student, is far

beyond the reach of AI's standardised feedback (e.g. "Good try!"). In terms of guiding thought, the teacher acts more like a cognitive alchemist: when a student offers a fragmented argument like "Remote learning. ... efficient... but lonel", a teacher can elevate the discussion by linking it to broader concepts, such as Plato's cave allegory. Such interventions foster cognitive leaps that transcend the mechanical associations of AI prompts.

4.2. Deep-Literacy in Cultural Contexts and Intercultural Pedagogy

Teachers facilitate cross-cultural understanding by moving beyond "explaining differences" to establishing "two-way connections". When explaining the concept of "polite refusal" in English, teachers not only compare the pragmatic differences between "Sorry, I'm tied up" and the Chinese "不好意思，我没空", but also guide students to reflect on the underlying cultural priorities, such as "collective harmony" versus "personal boundaries". This approach, which links linguistic phenomena to deeper cultural structures, enables students to grasp not merely isolated rules, but the underlying logic of values behind them. In practice, this involves establishing "local cultural anchors." For example, when discussing "Western table manners," a teacher might introduce China's "round-table culture" as a comparative lens. This bidirectional approach prevents the "one-way cultural input" typical of AI. This defines the teacher as a "cultural responder" and navigator, a role irreplaceable by any simplified, technological knowledge repository.

4.3. Responsive Regulation of Classroom Emergence

A teacher's ability to respond in real time to the emergent dynamics of a speaking classroom is a unique strength that generative AI cannot match. During spoken interaction, students' thinking frequently transcends the boundaries of predetermined topics, giving rise to creative expressions that may appear to stray from the subject—for instance, whilst discussing environmental protection measures, a student might suddenly draw a connection to the sustainable use of natural materials in local traditional crafts. Such leaps often reveal valuable interdisciplinary connections and personalised insight. At such moments, teachers can keenly recognise the value of such expressions and adjust the direction of the conversation accordingly by inviting the student to elaborate on the connection between traditional crafts and environmental principles. Such guidance not only respects the student's intellectual autonomy but also transforms this deviation into an opportunity to deepen the dialogue, shifting the oral classroom from a "scripted progression" to one that "grows with thought", thereby fostering a vibrant atmosphere of inquiry. In contrast, generative AI, constrained by pre-set algorithmic frameworks and topic scopes, often exhibits clear limitations when faced with such "deviations". When a student's expression goes beyond the pre-set logic in its database, the AI typically responds with phrases such as "Let's get back to the topic of environmental protection" or "This topic is not within the scope of this exercise",

forcibly steering the conversation back onto the predetermined track. Whilst this mechanical correction ensures standardisation of the teaching process, it risks stifling students' creative thinking and desire to express themselves. This discrepancy highlights a core limitation of technological standardization: it lacks the flexible pedagogical wisdom to productively accommodate uncertainty. Yet, it is precisely this wisdom in guiding emergent scenarios that is key to fostering students' cognitive flexibility and expressive individuality.

5. Operationalizing Boundaries: A Framework for Human-AI Collaboration

5.1. A Three-Dimensional Criteria Framework

A three-dimensional framework—encompassing task attributes, interaction depth, and cultural relevance—clarifies the division of roles between generative AI and teachers. In terms of task attributes, AI's strengths as an aid lie in handling quantifiable, highly repeatable, mechanical tasks—such as pronunciation training or sentence pattern practice in fixed scenarios. Such standardised operations can significantly enhance training efficiency by leveraging AI's ability to iterate on algorithms; teachers, however, must retain control over areas requiring dynamic adjustment and cultural sensitivity (Anderson & Krathwohl, 2001). This is particularly critical in open-ended cross-cultural dialogue, where teachers must decode contextual subtext (e.g., the refusal connoted by "I' ll think about it" in negotiations) and guide students to understand the value logic behind cultural symbols. The depth of interaction reveals a second benchmark for human-machine collaboration: AI can efficiently handle superficial information transfer, such as practising basic grammatical rules through pre-set question banks or automatically marking verb tenses, where the feedback mechanism is essentially a closed-loop data-matching process. However, when teaching enters the realm of deep interaction where emotion and thought intertwine, the humanistic value of the teacher becomes irreplaceable. This reflects the core role of the teacher as a 'scaffolder' within social constructivist theory (Vygotsky, 1978). For instance, in a debate on "Should AI replace human teachers?", teachers must analyze argument logic, identify cognitive blind spots, and—through techniques like strategic pauses—foster the psychological safety necessary for critical thinking. Such complex interactions require the teacher's pedagogical wisdom and emotional empathy, which completely transcend the boundaries of AI's algorithmic decision-making. Finally, the degree of cultural relevance determines the value threshold for the division of labour between humans and machines. Low-cultural-load linguistic forms (e.g., English tense rules) can be systematically presented by AI; indeed, their value-neutral nature actually ensures the precision of knowledge transfer; however, when dealing with expressions bearing high cultural load—including historically laden slang (e.g., 'spill the beans') or identity-carrying metaphors (e.g., "soul food" in

Black English)—the agency of the teacher as a cultural interpreter becomes crucial (Gay, 2018). Teachers must engage in cultural translation and deconstruct cultural hegemony (e.g., critiquing Western-centric narratives in teaching materials), thereby elevating language teaching to a practice of cultural dialogue. Together, these dimensions form a set of collaboration rules that harness AI's efficiency while safeguarding education's humanistic core.

5.2. A Four-Quadrant Task Taxonomy: The Case of Oral Instruction

Table 1 presents a four-quadrant model for human-machine collaboration in oral instruction, mapping tasks along the dimensions of Technical Efficacy and teacher irreplaceability. It operationalizes the TPACK framework by defining how technological, pedagogical, and content knowledge can be integrated (Mishra & Koehler, 2006). In the top-left quadrant (high technical efficacy × high teacher irreplaceability), pronunciation correction is a typical task. Leveraging advanced speech recognition algorithms, AI can accurately detect deviations at the phonemic level, such as distinguishing the spectral differences between /θ/ and /s/. However, diagnosing the root causes of errors—such as distinguishing cultural misuses resulting from negative transfer from the learner's first language—still requires teachers to take the lead, drawing on their professional knowledge and teaching experience, thereby establishing a virtuous cycle of human-machine complementarity. The top-right quadrant (low technical efficacy × high teacher irreplaceability) centralizes the teacher's unique role in deep cultural interpretation and in providing affective support through non-verbal feedback. The lower-left quadrant (high technical efficacy × low teacher irreplaceability) fully harnesses the engineering advantages of AI. Standardised simulation tasks, such as recreating airport check-in dialogue processes, can be entirely executed by AI systems, with teachers required only to conduct spot checks on quality as appropriate. Thus, the model clarifies the core logic of collaboration: technical efficacy dictates AI's feasibility, while teacher irreplaceability defines necessity. This delineation helps prevent role encroachment and guides the optimization of human-AI collaboration in speaking instruction.

Table 1. A Four-Quadrant Model for Human-AI Collaboration in Oral English Teaching

Teacher Irreplaceability Technical Efficacy	High ↑	Low ↓
High ↑	Pronunciation Correction (AI-Assisted)	Cultural Interpretation & Emotional Resonance (Teacher-Core)
Low ↓	Contextual Simulation (AI-Lead)	Redundant Zone (Inefficient Tasks)

5.3. Pathways for Mitigating Risks in Practice

Mitigating the risks of generative AI in spoken English instruction requires a dual-track approach: teacher development and tool redesign. At the teacher level, the core lies in delivering integrated training in "AI literacy and cultural teaching

competence". Such training transcends the limitations of mere technical operation, requiring teachers not only to be proficient in using AI tools to analyse spoken English training data, but also to cultivate critical thinking that goes beyond algorithmic constraints, with a particular focus on enhancing sensitivity to cultural subtext and emotional cues. Put simply, this requires teachers to develop a form of 'critical data literacy' that transcends technology, enabling them to interpret and utilise AI-generated data in a discerning manner (Mandinach & Gummer, 2016). When AI flags a particular student expression as "incorrect", teachers must consider the cultural context to determine whether it is valid within that specific context. This dual-track training shapes teachers into dynamic knowledge mediators who optimise teaching by leveraging the efficiency of AI whilst upholding their role as professional reflectors, thereby safeguarding the humanistic core of language teaching amidst technological application. At the level of tool design, innovation must respond to the principles of co-design, mitigating the risks of algorithmic bias through multi-stakeholder collaboration between teachers, students and technical experts. The focus lies in constructing customisable cultural modules, incorporating localised corpora (e.g., the conversational logic of honorifics used in Chinese family dinners) into AI training frameworks, thereby freeing generated content from the shackles of Western-centrism. This echoes the warnings of critical educational technology regarding algorithmic bias and puts into practice a pathway to promoting educational equity through diverse corpora and localised design (Benjamin, 2019). In practice, this involves allowing the customization of cultural parameters to ensure language output reflects authentic cross-cultural dynamics. This dual optimization forms a risk-mitigation mechanism that achieves a dynamic balance between teaching efficacy and cultural integrity.

6. Conclusion

This study has delimited the respective roles of generative AI and human teachers in spoken English instruction. Generative AI serves as an auxiliary tool for mechanical, low-culture-load tasks, while the teacher's irreplaceable role lies in affective, cultural, and adaptive cognitive domains. These tasks—rooted in human care, cultural nuance, and creative thought—form an essentially human domain, reliant on personal agency and cultural identity in ways that resist technological replication. The boundary is drawn not by efficiency, but by whether a task engages the core human elements of learning. This criterion provides the fundamental framework for human-machine collaboration, achieving a symbiotic balance. Looking ahead, collaboration should evolve towards deeper integration, exemplified by a closed-loop teaching process and, more critically, a "teacher-AI collaborative reflection" model. This bidirectional synergy enables teachers to refine instruction with AI-generated data while using their professional judgment to calibrate AI's cultural and pedagogical outputs, ensuring that AI evolves as an enabler of teaching, not a replacement for the teacher.

This bidirectional, collaborative model will ultimately strike a balance between technological efficiency and humanistic values, ensuring that generative AI truly serves as an enabler of high-quality spoken English instruction, rather than a replacement.

References

[1] Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition.* New York: Longman.

[2] Benjamin, R. (2019). *Race after technology: Abolitionist Tools for the New Jim Code.*Polity.

[3] Feng, Y. G., & Zhang, L. (2024). The spatial dimension of moral education in the era of generative artificial intelligence. *China Educational Technology*, (5), 45-52.

[4] Freire, P. (1970). *Pedagogy of the Oppressed*.New York: Continuum.

[5] Gay, G. (2018). *Culturally Responsive Teaching: Theory, Research, and Practice* (3rd ed.).Teachers College Press.

[6] Gu, X. W. (2018, April 27). Technology as both antidote and poison: A dialogue with French philosopher Bernard Stiegler. *Jiefang Daily*, p. 09.

[7] Hou, H. X., & Wang, D. (2025). Risks, concerns and avoidance paths of teachers' teaching innovation in the era of generative artificial intelligence. *China Educational Technology*, (3), 20-26.

[8] Huang, L., Baker, J., & Reddy, R. (2022). A historical perspective of speech recognition. *Communications of the ACM,* 57(1), 94-103.

[9] Mandinach, E.B., & Gummer, E.S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, 60, 366-376.

[10] Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record,* 108（6）, 1017-1054.

[11] Richards, J. C. (2006). *Communicative Language Teaching Today*. Cambridge University Press.

[12] Rogers, C. R. (1969). *Freedom to Learn: A View of What Education Might Become.* Columbus, OH: Merril Pub. Co.

[13] Stiegler, B. (1998). *Technics and Time, 1: The Fault of Epimetheus.* Stanford University Press.

[14] Stiegler, B. (2013). *What Makes Life Worth Living: On Pharmacology*. Polity Press.

[15] Vygotsky, L. S. (1978).*Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.