Modern Linguistics 现代语言学, 2026, 14(11), 245-256
Published Online November 2026 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2026.14111069

# 生成式人工智能影响二语口语表现的研究综述

**张　迪，陈秀雅**

大连海事大学外国语学院，辽宁 大连

收稿日期：2026年9月30日；录用日期：2026年10月29日；发布日期：2026年11月11日

## 摘　要

**随着ChatGPT等生成式人工智能(GenAI)的迅速普及，其在二语口语教学中的应用备受学界关注，但已有研究多将口语与其他语言技能合并讨论，对GenAI融入口语练习的路径及对口语表现的具体影响缺乏系统梳理。本研究依据PRISMA 2020规范，对ChatGPT发布以来的相关研究进行系统综述，最终纳入33篇文献。结果表明，现有研究多采用通用型与任务导向型工具，以亚洲高校英语学习者为主要对象，方法以量化与混合研究为主；GenAI可促进学习者整体口语表现及各维度提升，但仍面临技术故障、反馈质量不稳、人际互动不足等挑战。本研究厘清了该领域研究图景，为后续研究与教学实践提供参考。**

## 关键词

**生成式人工智能，二语口语表现，系统综述，口语教学，人机互动**

# A Review of Research on the Impact of Generative Artificial Intelligence on L2 Speaking Performance

**Di Zhang, Xiuya Chen**

School of Foreign Languages, Dalian Maritime University, Dalian Liaoning

Received: September 30, 2026; accepted: October 29, 2026; published: November 11, 2026

## Abstract

**With the rapid proliferation of generative artificial intelligence (GenAI) tools such as ChatGPT, their application in second language (L2) speaking instruction has garnered growing scholarly attention. However, existing studies have mostly examined speaking skills alongside other language skills, leaving a lack of systematic synthesis regarding how GenAI integrates into L2 speaking practice and its specific impacts on speaking performance. Following the PRISMA 2020 guidelines, this study**

文章引用: 张迪, 陈秀雅. 生成式人工智能影响二语口语表现的研究综述[J]. 现代语言学, 2026, 14(11): 245-256.
DOI: 10.12677/ml.2026.14111069

**presents a systematic review of research published since the launch of ChatGPT, with 33 Chinese and English academic articles included in the final sample. The findings show that existing studies predominantly employ general-purpose and task-oriented GenAI tools, target tertiary-level English learners in Asia, and adopt mainly quantitative and mixed-methods designs. While GenAI facilitates improvements in learners' overall speaking performance and its sub-dimensions, its application is constrained by technical glitches, inconsistent feedback quality, and insufficient interpersonal interaction. This study maps the research landscape of GenAI in L2 speaking instruction and offers implications for future research and pedagogical practice.**

## Keywords

**Generative Artificial Intelligence, L2 Speaking Performance, Systematic Review, Speaking Instruction, Human-AI Interaction**

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

## 1. 引言

### 1.1. 研究缘起

2022 年 11 月，OpenAI 发布 ChatGPT，生成式人工智能由此迅速进入公众与教育研究者的视野。以大语言模型(Large Language Models, LLMs)为核心的 GenAI 凭借开放域对话、上下文理解与多轮交互等能力，引发了教育领域对教学支持、学习反馈与人机协作模式的广泛讨论[1]。Lo [2]通过快速文献综述指出，ChatGPT 对教育的影响已涵盖教学设计、学习评估与学术诚信等多个维度，其渗透速度之快促使各国教育政策制定者不得不正视这一技术变革。UNESCO [3]发布的《生成式人工智能教育与研究应用指南》亦从政策层面确认了 GenAI 已成为全球教育领域不可忽视的重要议题。

在语言教育领域，生成式人工智能为学习者提供了区别于传统课堂的练习空间：它能够根据上下文生成较为自然、连贯的回应，并支持追问、澄清、角色扮演和反馈修正等多种互动形式，因而在语言练习、口语输出和交际模拟等方面表现出较强的应用潜力，迅速成为外语教育研究中的热点议题。

口语能力作为二语习得的核心技能之一，直接关系到学习者的交际效能与语言综合运用水平。Bygate [4]在其经典著作中指出，口语交际涉及语言形式、互动管理与意义协商等多重认知加工过程，是语言能力中最复杂、最难精通的维度之一。然而，传统口语教学长期面临若干现实限制：课堂时长有限，学习者个体化练习机会不足；教师难以及时向每位学习者提供持续且有针对性的反馈；同时，公开展示式课堂互动也可能提升学习者的外语焦虑水平[5]。Fathi 等[6]进一步指出，在传统教学环境中，学习者因担心负面评价而回避口语输出，形成“焦虑–回避–能力停滞”的恶性循环。GenAI 所具备的即时交互、较高可及性与个性化反馈等特征，为破解上述困境提供了新的技术路径。随着相关应用不断增加，系统梳理生成式人工智能在二语口语学习中的融入方式及其实际影响，已成为当前外语教育研究中具有现实意义的重要任务。

### 1.2. 研究空白与目的

近年来，人工智能辅助语言学习的有效性已得到大量实证研究的支持，相关综述与元分析亦不断涌现。已有元分析与系统综述总体上证实了 AI 技术对外语学习的促进作用[7]-[10]，但在生成式人工智能辅

助口语学习这一新兴议题上仍存在明显不足：其一，不少研究将口语与阅读、写作、词汇等技能合并分析，较少单独呈现口语这一产出性技能的具体表现及其变化特征；其二，部分研究将传统基于规则的聊天机器人与 GenAI 工具一并纳入讨论，尚未充分揭示不同技术类型在交互深度、反馈方式和任务支持上的差异；其三，对 GenAI 究竟以何种工具形态、在何种教育情境中融入口语练习，及其对口语表现各维度的具体影响，仍缺乏较为系统的归纳。

基于上述背景，本研究聚焦 ChatGPT 发布至今生成式人工智能辅助二语口语学习的相关研究，尝试回答以下两个问题：第一，GenAI 工具如何融入口语练习？第二，生成式人工智能如何影响口语表现？

本文希望据此系统呈现该领域的研究图景，梳理生成式人工智能在口语教学与学习中的应用路径、主要成效与现实挑战，从而为后续研究与教学实践提供参考。

### 1.3. 论文结构

除引言外，本文共包含四个部分。第二部分为研究背景，阐述核心概念、理论基础与技术发展脉络，并梳理研究现状与空白；第三部分为研究结果与讨论，介绍文献检索筛选流程并围绕两个研究问题呈现主要研究发现；第四部分为研究不足与未来展望，分析现有研究的局限性并提出未来研究的改进方向；最后为结语，归纳全文主要发现并强调本研究的核心观点。

## 2. 研究背景

### 2.1. 核心概念与理论基础

#### 2.1.1. 生成式人工智能的界定

生成式人工智能是指基于大语言模型、通过深度学习技术自主生成自然语言内容的新型人工智能系统。在教育领域，以 ChatGPT 为代表的 GenAI 聊天机器人依托 Transformer 架构与海量语料训练，具备开放域自然语言理解、长文本关联推理与多轮连贯交互等核心能力[11]。这类模型可以动态理解交际语境，实时生成全新文本，而非简单调取数据库内已有内容，为二语口语练习提供了全新交互载体。与传统基于规则或检索式的聊天机器人相比，GenAI 的输出不再完全依赖预设脚本，而是能够根据对话上下文生成较为个性化的回应，在应答的开放性、语境适应性与连续交互能力方面表现出更强的灵活性，能够模拟真实口语交际场景，对学习者的二语口语产出产生潜在影响。

#### 2.1.2. 二语口语表现的测量维度

二语口语表现的评估需兼顾认知产出与情感参与两个层面。在认知层面，本研究采用 Skehan [12]提出的 CAF 三元分析框架，即以复杂度(complexity)、准确度(accuracy)和流利度(fluency)作为核心观测维度，分别衡量句法结构难度、语言形式正确程度与产出的连贯性；同时结合 Skehan [12]将词汇(lexis)纳入口语表现建模的思路，关注学习者在词汇层面的表达资源与变化程度。Luoma [13]进一步指出，口语评价不应局限于语言形式的正确性，还应关注学习者在真实交际情境中的语言运用能力与互动策略。鉴于口语交际具有实时性与互动性，本研究在 CAF 框架基础上进一步关注发音(pronunciation)维度的改善，其与口语交际的可理解性直接相关。在评价标准上，欧洲语言共同参考框架(CEFR)从语言范围、准确性、流利度、互动能力与连贯性等方面描述口语能力，为跨研究的比较提供了参照；然而当前实证研究的测量指标与 CEFR 的对接仍不充分，测量工具不尽统一，这也是本综述关注的方法论问题之一。

#### 2.1.3. 核心理论基础

本研究的理论框架建立在三个经典二语习得假说之上，分别从输出驱动、互动机制与情感调节三个维度解释 GenAI 辅助口语学习的可能机理。首先是 Swain [14]的输出假说(Output Hypothesis)，该假说认

为可理解输出(comprehensible output)具有注意触发、假设检验与元语言反思三重功能，能够推动学习者注意到自身语言系统与目标语之间的差距并修正语言假设。其次是 Long [15]的互动假说(Interaction Hypothesis)，该假说强调意义协商(negotiation of meaning)与修正性反馈(corrective feedback)对二语发展的促进作用，对话双方通过确认检查、澄清请求等策略调整输入与输出，既提高可理解性，又引导学习者关注语言形式。再次是 Krashen [16]的情感过滤假说(Affective Filter Hypothesis)，该假说主张低焦虑、高动机的情感环境能够降低“情感过滤”，促进可理解输入的有效吸收。GenAI 所提供的即时对话、反馈回应与低评价压力环境，为在数字技术情境下探讨上述三个假说的适用性提供了新的经验场景，也为本研究分析生成式人工智能影响口语表现的内在逻辑提供了理论视角。

除上述三大假说外，社会文化理论(Sociocultural Theory)为理解 GenAI 的支架功能、回答“GenAI 工具如何融入口语练习”(研究问题 1)提供了补充视角。Vygotsky [17]提出的最近发展区(Zone of Proximal Development, ZPD)概念强调，学习者在更有能力的他者支架支持下能够完成独立无法完成的任务；Lantolf 和 Thorne [18]将这一理论应用于二语习得领域，指出对话互动本身就是一种中介工具。GenAI 通过动态调整语言难度、提供即时提示与个性化反馈，在一定程度上扮演了“数字支架”的角色，与上述理论形成呼应。

## 2.2. 技术辅助口语学习的发展与生成式 AI 的介入

### 2.2.1. 技术辅助口语学习的发展脉络

计算机辅助语言学习(Computer-Assisted Language Learning, CALL)自 20 世纪 60 年代兴起以来，经历了从行为主义到交际取向再到整合取向的范式演进。Warschauer 和 Healey [19]将 CALL 的发展划分为三个阶段：早期的行为主义 CALL 侧重于通过计算机进行反复的语法与词汇操练；20 世纪 80 至 90 年代的交际性 CALL 强调利用技术创设语言使用情境，促进学习者之间的真实交流；进入 21 世纪后，整合性 CALL 则致力于将多媒体、互联网与语言学习过程深度融合，实现技能的综合训练。随着移动设备的普及，移动辅助语言学习(Mobile-Assisted Language Learning, MALL)进一步突破了时空限制，使口语练习从固定的计算机实验室延伸至日常碎片化场景。

在技术辅助口语学习的具体实践中，基于对话的计算机辅助语言学习(Dialogue-based CALL)逐渐成为一个独立的研究领域。Bibauw 等[20]通过系统综述指出，对话式 CALL 涵盖智能计算机辅助语言学习(ICALL)、计算机辅助发音训练(CAPT)、口语对话系统(SDS)及聊天机器人(Chatbots)等多种研究流派，其共同特征在于通过人机对话为学习者提供口语产出与互动协商的机会。其中，计算机辅助发音训练及自动语音识别(ASR)技术的引入，使口语练习从单纯的文本互动拓展至发音层面的即时反馈[21]。然而，早期对话系统多基于规则匹配或检索式应答，交互灵活性有限，难以应对开放域的口语交际需求。进入 21 世纪第二个十年，基于人工智能的神经网络已展现出显著的语言处理能力[22]，为后续 GenAI 进入语言学习场景奠定了技术基础。

### 2.2.2. 生成式 AI 的介入及其潜在优势

2022 年 11 月 ChatGPT 的发布，被普遍视为 GenAI 加速进入语言学习场景的重要节点，以大语言模型为核心的工具在口语学习领域展现出传统技术难以比拟的潜在优势。首先，GenAI 能够为学习者营造相对低焦虑、较少评判压力的练习环境：相较于面对面的课堂展示和同伴互动，人机对话通常不会直接带来明显的社会评价压力，这使部分学习者更愿意尝试表达和反复练习。Fathi 等[6]的准实验研究发现，与面对面同伴互动相比，使用 AI 聊天机器人进行口语练习的学习者在流利度、词汇运用、语法准确性及发音等方面均有更显著的提升，同时其外语口语焦虑显著降低、交际意愿明显增强。其次，GenAI 通常能够提供较为即时、个性化的反馈支持，学习者在与 AI 的互动中可以获得语法、词汇、表达方式乃至任务完成情况方面

的即时回应，这种高频率的人机“微互动”在一定程度上接近 Long [15]所强调的意义协商过程。

再次，GenAI 支持情境化对话与任务模拟，学习者可根据自身需求设定日常交流、旅游问询、学术讨论或求职面试等不同场景，在较贴近真实交际的语境中练习表达、调整语言策略并积累话语资源。此外，GenAI 具有较高可及性，在一定程度上突破了传统课堂的时间限制，为学习者开展持续性、个体化的口语练习提供了更多可能。无论是在课内作为辅助工具，还是在课外作为自主练习资源，生成式人工智能都表现出较强的延展性与适配性，这也使得“GenAI 工具如何融入口语练习”成为值得专门考察的重要问题。

## 2.3. 研究现状与空白

### 2.3.1. 已有综述与元分析的贡献

近年来，人工智能辅助语言学习领域的综述与元分析研究呈快速增长态势，为理解技术介入语言教学的整体效应提供了重要证据。Huang 等[10]对聊天机器人在语言学习中效用的系统综述表明，聊天机器人能够在提升学习者动机、降低焦虑及促进语言技能发展方面发挥积极作用，但其效果受技术类型、任务设计及学习者特征等因素影响。Zhang 等[9]通过元分析进一步量化了聊天机器人辅助语言学习的整体效应，纳入 18 项实验研究的结果显示，该技术对语言学习具有中等程度的积极影响(Hedges’ g = 0.527)，且在 K-12 教育阶段与词汇学习领域效果尤为突出。王乐等[7]与 Xu 等[8]的元分析研究则分别整合 48 项与 15 项实证研究，从更宏观的视角证实了人工智能对外语学习具有中等至大的正向效应，并识别出教育层次、语言技能、技术类型与干预时长等关键影响因素。

与此同时，随着 ChatGPT 等生成式人工智能工具的迅速普及，相关研究开始从更聚焦的角度讨论大语言模型在语言教学中的应用。Lo [2]和 Kohnke 等[11]分别从教育影响与语言教学应用出发，指出 GenAI 在教学设计、反馈支持、任务构建和互动练习等方面具有新的可能性，为理解生成式人工智能进入语言教育场景提供了重要的理论和实践启发。

### 2.3.2. 已有研究的不足

尽管上述综述与元分析贡献显著，但针对生成式人工智能与二语口语表现的专门性审视仍存在明显缺口，主要体现在以下五个方面。

第一，现有研究往往将口语与听、读、写、词汇等语言技能合并讨论，较少聚焦生成式人工智能对口语表现的具体影响，弱化了口语作为产出性技能所具有的实时性、互动性与多维评价特征。第二，部分研究对不同类型 AI 工具的区分仍不够充分，常将传统聊天机器人与生成式人工智能置于同一框架下讨论；Derakhshan 和 Ghiasvand [23]指出，ChatGPT 等生成式人工智能在自主性、语境适应性与开放域交互方面明显超出传统聊天机器人，有必要对这一技术代际变化进行更细致的梳理。第三，现有研究在 GenAI 工具的具体融入方式方面仍缺乏系统归纳，生成式人工智能可能作为课堂辅助工具、课外自主练习平台、任务对话伙伴、角色扮演对象或反馈修正支持系统出现，但这些实践路径分散在个别实证研究中，尚未形成清晰的整合图景。第四，口语表现的测量维度与评价标准尚未形成较为统一的框架，不同研究采用的指标从整体口语成绩到流利度、准确性、复杂度、发音、词汇和语法各不相同，影响了研究结果之间的比较与整合。第五，尽管不少研究报告了生成式人工智能对口语学习的积极作用，但其应用过程中的限制与挑战尚未得到充分整合，如技术故障、语音识别偏差、内容准确性不足以及人际互动和情感支持缺失，而这些同样是理解其口语应用图景的重要组成部分。

### 2.3.3. 本综述的研究定位

基于上述研究空白，本综述聚焦 ChatGPT 发布至今生成式人工智能辅助二语口语学习的相关研究，

系统梳理该领域的发展图景。与已有偏重宏观整合的综述与元分析不同，本文将研究视角集中于口语这一具有实时性、互动性和多维评价特征的语言技能，围绕两个核心问题展开：其一，GenAI 工具如何融入口语练习，考察生成式人工智能在不同教育情境中的应用形态；其二，生成式人工智能如何影响口语表现，以及相关研究中呈现出的应用困难与现实挑战。

通过上述分析，本文希望为生成式人工智能辅助口语教学提供更清晰的研究图景与更具针对性的理论参照，也为未来在英语及更广泛二语教育场景中的实践探索提供参考。

## 3. 研究结果与讨论

本研究文献检索与筛选过程严格遵循 PRISMA 2020 [24]相关要求，选取 Web of Science、Scopus 以及 ERIC 等三个数据库，使用检索词 GenAI OR “generative AI” OR chatbot OR “large language model” OR “LLM” AND EFL OR ESL OR English AND “speaking performance” OR “speaking outcome” OR “oral proficiency” OR pronunciation OR fluency OR accuracy。本研究依据预先确定的纳入与排除标准对检索所得文献进行筛选，具体标准见表 1。纳入文献须为 2022~2026 年发表于同行评议学术期刊的实证研究，采用生成式人工智能技术开展英语口语学习或教学干预，收集数据包含口语表现的前测与后测，并报告可用于评价口语表现变化的量化结果。测量内容可以涉及整体口语能力、流利度、准确性、复杂度、发音等具体维度，或同时涵盖整体口语能力与具体维度。相应地，本研究排除了不符合发表时间及文献类型要求的研究，以及未使用生成式人工智能辅助口语干预，或者未明确指出人工智能类型的、缺少口语表现前测或后测或未报告口语表现量化结果的研究。值得注意的是，在筛选文献时，考虑到不同研究的干预形式与实施周期存在差异，本研究未对干预时长设置统一的最低限制，而是在数据提取阶段对各研究的实际干预时长进行记录与编码，并在后续综合分析中予以考虑。

**Table 1.** Inclusion and exclusion criteria
**表 1.** 纳入与排除标准

| 纳入标准 | 排除标准 |
|---|---|
| 发表时间为 2022~2026 年 | 发表时间不在 2022~2026 年范围内 |
| 发表于经过同行评议的学术期刊 | 会议论文、学位论文、综述性论文及其它未经同行评议的文献 |
| 研究对象为英语作为外语或第二语言的学习者 | 研究对象不是英语作为外语或第二语言学习者或研究对象不明确的文献 |
| 应用生成式人工智能技术进行干预 | 使用其他技术工具或未明确指出应用生成式人工智能进行干预 |
| 对英语口语表现或其具体维度进行量化测量 | 未提供口语表现或其具体维度量化测量结果 |

经过以上检索与筛选，本研究最终纳入文献共 33 篇并依据研究问题对最终纳入文献进行编码，编码表设置参考了过往相关文献的编码[25] [26]对纳入文献的基本信息、研究对象、研究设计及研究结果进行系统整理。编码内容主要包括作者、年份、国家、年龄、教育水平、英语水平、样本量、研究变量、研究方法、人工智能工具、理论基础、测试工具、主要研究结果及干预时长，以便对现有研究的总体特征与发展趋势进行比较分析。本研究由两名研究者依据预先制定的编码框架对纳入文献进行独立编码，并采用百分比一致率检验编码者间一致性。结果显示，两名编码者的一致率为 90%。为进一步保证编码过程的一致性，两名研究者在正式编码前对编码标准及各变量的操作性定义进行了统一。对于原文未明确报告的信息，相应编码项保持空白，以避免基于主观判断进行推断或补充。对于存在分歧的编码项目，通

过重新核对原文并协商讨论后达成一致。表 2 选取部分纳入文献作为编码示例，展示主要研究特征及相关编码信息，其余编码内容，包括学习者英语水平、研究变量、研究方法、理论基础及主要研究结果，则结合研究问题在后文进行综合分析。

**Table 2.** An illustrative sample of the coding scheme
**表 2.** 编码表部分内容示例

| 序号 | 作者 | 年份 | 国家 | 年龄 | 教育水平 | 样本量 | AI 工具 | 测试 | 时长 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Tai | 2026 | 中国 | 16~17 | 高中 | 33 | ChatGPT | GEPT | 12 周 |
| 2 | Tai & Chen | 2024 | 中国 | 12~13 | 小学 | 85 | CoolE Bot | GEPT | 3 周 |
| 3 | Rad | 2024 | 伊朗 | 20 | 大学 | 66 | Speeko | 雅思口语 | 10 周 |

## 3.1. 生成式人工智能工具种类

根据 GenAI 工具的开发目的与应用范围，本研究将相关工具划分为通用型 GenAI 工具和任务导向型 GenAI 工具。前者并非专门针对语言学习开发，而后者主要围绕特定语言学习目标或教学任务进行设计。前者通常具有更强的开放式对话能力和多场景适用性，后者则更强调围绕特定练习目标进行结构化任务支持。无论工具形态如何，GenAI 能够通过持续对话、表达修正、情境模拟与反馈提示等方式，为学习者提供较为灵活的口语练习支持，因此在二语口语教学与学习中受到广泛关注。

### 3.1.1. 通用型 GenAI 工具

在 GenAI 工具的选择上，现有研究多采用功能广泛、开发较为成熟的通用型 GenAI 工具作为干预载体[27]，主要包括面向大众的通用型聊天机器人(如 ChatGPT、豆包)和由大语言模型驱动的智能语音助手(如 Google Assistant、Siri)。例如，Behforouz 等[28]探究了移动端 ChatGPT-4o 对 62 名中级水平 EFL 大学生口语能力的影响。在为期一个月的干预中，使用 ChatGPT-4o 进行课外互动的实验组，其整体口语表现以及语法准确性、任务回应度等维度的提升均显著优于控制组。

总体而言，通用型 GenAI 工具能够依托即时交互、反馈及多模态功能，为学习者提供较为灵活的口语练习支持，并对其口语表现产生积极影响。

### 3.1.2. 任务导向型 GenAI 工具

除通用型 GenAI 工具外，部分研究采用针对特定语言学习任务或教学情境开发的任务导向型 GenAI 工具。Hsu 等[29]考察了任务导向型聊天机器人 TPBOT 对 48 名低分段大学生口语水平的影响。在为期 4 个月的实验中，使用 TPBOT 进行课外互动与即时纠偏的实验组，其口语提升幅度显著高于使用传统教材的控制组，且练习频次与成绩增幅呈正向关联。

总体而言，从社会文化理论视角来看，两类工具均可能通过提供语言提示、互动反馈等方式发挥支架作用，但其支持形式存在差异。通用型 GenAI 工具能够根据学习者的即时输入灵活调整互动内容，更接近开放式支架；而任务导向型工具则通过预设任务和针对性反馈，为特定口语技能的练习提供较为结构化的支持。这些功能在一定程度上体现了最近发展区理论所强调的外部支持对学习者能力发展的促进作用。

## 3.2. 教育情境与研究方法

本节主要针对生成式人工智能影响学习者口语表现相关研究的教育情境与研究方法进行总结。总体来说，相关研究更多关注亚洲国家大学阶段的英语学习者，研究方法则以量化或混合方法为主。

#### 3.2.1. 教育情境

在本研究纳入的 33 篇文献中，研究对象多为大学阶段的英语学习者[30] [31]，另有部分研究关注生成式人工智能对青少年英语学习者口语能力的影响[32]。举例来说，Yang 等人[33]以 314 名不同英语水平的 10~15 岁中小学生为研究对象，考察其在真实课堂中开展生活场景口语任务的表现。

就已报告相关信息的文献来看，研究参与者年龄范围在 10~25 岁之间，英语水平主要集中在 B1~B2 之间，多数参与者来自亚洲国家，主要集中在中国，此外还有日本、韩国、马来西亚、伊朗、沙特阿拉伯、土耳其等。相比之下，其他地区国家开展的有关人工智能对口语能力影响的研究数量较少，例如，Sayed 等人[34]以埃塞俄比亚高校 EFL 学习者为对象，考察了生成式人工智能对英语口语表现的影响，结果显示其对学习者口语能力具有积极促进作用。

#### 3.2.2. 研究方法

在研究设计方面，多数研究为混合研究，少部分为量化研究[35]。其中量化研究多采用前测加后测的设计用以收集参与者口语表现分数，通过对两次收集数据的分析判断生成式人工智能对口语能力变化是否存在影响。质性研究主要通过访谈、学习反思等深入了解学习者对于生成式人工智能参与口语学习的态度以及学习者在参与实验期间的动态变化[36]。例如，吴坚豪等[37]采用实验组与控制组前后测设计，以 100 名大学英语学习者为对象开展为期 16 周的教学实验，通过比较两组在口语复杂度、准确度和流利度等指标上的变化考察生成式人工智能的干预效果。Wang 等[38]则采用问卷与访谈相结合的混合研究设计，以 240 名中国高校 EFL 学习者为研究对象，在量化考察口语流利度与准确性变化的同时，通过访谈收集学习者对 ChatGPT 使用体验、学习自主性及发音指导等方面的反馈。

对比两类研究设计可见，量化研究侧重于通过标准化测试等数据客观检验口语能力的干预效果与显著性；而混合研究则在量化的基础上，进一步通过质性文本补充了对情感态度、学习自主性及技术局限的深层解释，从而弥补了单一量性设计难以捕捉主观体验与内在原因的局限。

### 3.3. 生成式人工智能对口语能力影响的实证效果

本节主要分析生成式人工智能对英语学习者口语能力的实证影响。总体来看，现有研究普遍显示生成式人工智能对学习者口语能力具有积极作用，但其影响因评价指标不同而呈现一定差异。部分研究以整体口语成绩衡量干预效果，另一些研究则进一步考察流利度、准确性、发音和复杂度等具体维度。

#### 3.3.1. 整体干预效果

在以整体口语能力作为衡量指标的研究中，生成式人工智能总体呈现出积极的干预效果。相关研究采用的测量工具包括研究者或教师设计的口语测试[39] [40]，以及托业[41]、雅思[42]等标准化语言测试。例如，Lee [43]采用多媒体辅助英语口语测试(Multimedia Assisted Test of English Speaking, MATEST)，考察了 GenAI 与同伴混合练习对 39 名韩国 EFL 大学生口语产出的影响。经过 14 周的干预，学习者后测口语成绩较前测显著提高。总体而言，以综合测试成绩为评价指标的研究较为一致地表明，GenAI 辅助练习能够促进学习者整体口语表现的提升。

#### 3.3.2. 口语能力具体维度

除整体口语成绩外，部分研究进一步从不同维度考察生成式人工智能对口语能力的影响。参照 CEFR 的口语能力维度划分，口语能力可从语言范围、准确性、流利度、互动能力和连贯性等方面进行评价。在纳入研究中，流利度和准确性是较为常见的评价指标[32]，部分研究还考察了发音和复杂度等维度[44]。例如，开展于 2025 年的一项研究[45]从复杂度、准确度和流利度三个维度比较学习者与 AI 同伴和人类同伴完成词语猜测任务时的口语表现。结果显示，与 GenAI 互动后，学习者的复杂度提升最为明显，流

利度也有所提高，而准确度的增幅与人类互动条件相近。Ye 等[46]则从语法和发音两个方面考察聊天机器人对 EFL 学习者口语准确性的影响，结果发现，学习者的语法和发音均显著改善，多类语法错误以及音位、删音和增音错误有所减少，但重音错误的改善相对有限。

综合来看，GenAI 辅助口语学习的积极作用并非均衡体现在各个口语维度上。现有研究显示，学习者在流利度、准确性、复杂度和发音等方面均可能获得改善，但不同维度的提升程度存在差异，且同一维度内部的不同指标也可能表现出不同的干预效果。因此，相较于笼统地认为 GenAI 能够提升口语能力，现有证据更表明其作用具有一定的维度差异。复杂度提升较明显，可能与输出假说强调的假设检验和元语言反思有关，而低评价压力为此提供了条件；准确度提升相对有限，可能与 GenAI 反馈“意义优先、形式聚焦不足”有关；流利度和发音改善，则可能与情感过滤降低、开口意愿和练习频次提升相呼应。

## 3.4. 生成式人工智能辅助口语能力提升的困难与挑战

为更全面地了解当前生成式人工智能在口语教学方面的局限与不足，本节主要梳理纳入文献中混合研究质性部分中参与者对于人工智能现存局限性的看法。

### 3.4.1. 技术运行故障

通过对纳入文献中质性部分进行整理分析，本研究发现参与者在使用 GenAI 工具进行口语练习时可能遭遇系统故障、网页异常及交互中断等技术问题。Ericsson 和 Johansson [47]在考察中学生使用对话式 AI 进行英语口语练习的过程中发现，技术故障可能影响学习者的使用体验。研究期间，AI 系统曾出现技术问题和系统漏洞，例如网页反复出现异常并中断学习者的正常使用，导致学习者产生明显的负面体验。

### 3.4.2. 内容与反馈问题

此外，生成式人工智能在辅助英语学习者进行口语练习时还会出现语音识别不准确、回复内容缺乏连贯性、反馈质量不高等问题[48]。Fathi 等[6]的质性分析显示，部分学习者指出 AI 聊天机器人存在内容准确性方面的局限，有时会产生错误或与对话内容不相关的回应，也可能无法准确理解学习者的输入。这表明，GenAI 生成内容的准确性与相关性仍是其应用于语言学习时需要关注的问题。从互动假说的视角来看，有效的意义协商和修正性反馈是促进二语发展的重要机制，而 GenAI 对学习者输入的错误理解及不准确反馈可能削弱互动的有效性，进而影响其对口语学习的支持作用。

### 3.4.3. 人际互动与情感缺失

部分参与者在使用生成式人工智能进行口语练习或者与聊天机器人对话时会感到 GenAI 无法提供与教师或同伴互动时获得的互动感以及情感联结[41]，Shafiee Rad [49]通过访谈考察学习者对 GenAI 辅助口语教学的看法。质性分析发现，尽管学习者认可 GenAI 提供的个性化反馈和便利性，但部分学习者认为 GenAI 辅助教学存在人际互动不足的问题，表示在学习过程中仍会怀念来自真实教师的指导与鼓励，这说明人工智能尽管能为英语学习者提供便利的口语练习和修正机会，显著提高学习者口语能力，但因其缺乏社交属性，无法在口语教学与学习中完全替代真实对话。这一发现与社会文化理论和情感过滤假说相呼应，社会文化理论强调社会互动与中介支持在语言学习中的重要作用，而情感过滤假说则关注焦虑、动机等情感因素对语言习得的影响。尽管 GenAI 能够提供即时反馈与个性化练习支持，但其在人际情感回应和真实社会交往方面仍存在一定局限，口语学习兼具认知与社交双重属性，人机互动难以完全替代真实人际互动的情感支持与社会协商功能。因此，GenAI 更适合作为传统口语教学的辅助工具，与教师指导和同伴互动形成互补。

## 4. 现有研究不足及未来发展方向

本研究仍存在一定局限。首先，受数据库选择、检索词及文献发表时间范围的限制，可能存在相关研究未被纳入的情况。未来研究可进一步拓展数据库和检索范围，并持续纳入最新研究成果，以提高研究覆盖度。其次，纳入文献在研究对象、GenAI 工具、研究设计及口语能力评价指标等方面存在较大差异，限制了不同研究结果之间的直接比较。未来可进一步细化不同学习者群体、GenAI 工具类型及口语能力维度，以更加准确地揭示生成式人工智能的具体作用。最后，本研究采用系统综述方法，结果以叙述性方式整合，尚未对干预效果进行量化整合。未来可在扩大研究样本的基础上开展元分析，进一步考察其对整体口语能力及不同口语维度的效应量，从而为生成式人工智能辅助英语口语教学提供更加可靠的实证依据。

## 5. 结语

总体而言，现有实证研究表明，生成式人工智能在促进英语学习者口语表现方面具有积极作用，但其干预效果呈现出一定的差异性。相较于以往覆盖多语言技能的宏观综述，本研究进一步明确：GenAI 对口语表现的促进作用存在维度异质性，且不同类型工具的支架功能各有侧重，这为后续精准化的教学应用提供了更具体的参照。具体而言，整体口语能力的提升较为一致，而流利度、准确性、发音和复杂度等具体维度的改善程度并不完全相同，表明生成式人工智能对口语表现的促进作用并非均衡分布于各个维度。这种差异提示，生成式人工智能的教学效果不仅取决于技术本身，还可能受到工具功能、练习任务及反馈方式等因素的影响。同时，技术运行不稳定、反馈内容准确性不足以及真实人际互动有限等问题，仍可能影响其教学效果。因此，未来研究应进一步关注不同生成式人工智能工具与具体口语能力维度之间的适配关系，并结合学习者特征、教学情境及干预周期，探索其发挥作用的条件与机制，从而为生成式人工智能在英语口语教学中的有效应用提供更加具体的实证依据。

## 参考文献


[1] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., *et al*. (2023) ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, **103**, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

[2] Lo, C.K. (2023) What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, **13**, Article 410. https://doi.org/10.3390/educsci13040410

[3] UNESCO (2023) Guidance for Generative AI in Education and Research. UNESCO.

[4] Bygate, M. (1987) Speaking. Oxford University Press.

[5] Horwitz, E.K., Horwitz, M.B. and Cope, J. (1986) Foreign Language Classroom Anxiety. *The Modern Language Journal*, **70**, 125-132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

[6] Fathi, J., Rahimi, M. and Derakhshan, A. (2024) Improving EFL Learners' Speaking Skills and Willingness to Communicate via Artificial Intelligence-Mediated Interactions. *System*, **121**, Article 103254. https://doi.org/10.1016/j.system.2024.103254

[7] 王乐, 刘甜, 赵波. 人工智能对外语学习成效影响的元分析[J]. 外语研究, 2026, 43(3): 84-91.

[8] Xu, G., Yu, A. and Liu, L. (2025) A Meta-Analysis Examining AI-Assisted L2 Learning. *International Review of Applied Linguistics in Language Teaching*, **64**, 905-937. https://doi.org/10.1515/iral-2024-0213

[9] Zhang, S., Shan, C., Lee, J.S.Y., Che, S. and Kim, J.H. (2023) Effect of Chatbot-Assisted Language Learning: A Meta-analysis. *Education and Information Technologies*, **28**, 15223-15243. https://doi.org/10.1007/s10639-023-11805-6

[10] Huang, W., Hew, K.F. and Fryer, L.K. (2022) Chatbots for Language Learning-Are They Really Useful? A Systematic Review. *British Journal of Educational Technology*, **53**, 310-329.

[11] Kohnke, L., Moorhouse, B.L. and Zou, D. (2023) ChatGPT for Language Teaching and Learning. *RELC Journal*, **54**, 537-550. https://doi.org/10.1177/00336882231162868

[12] Skehan, P. (2009) Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis.

*Applied Linguistics*, **30**, 510-532. https://doi.org/10.1093/applin/amp047

[13] Luoma, S. (2004) Assessing Speaking. Cambridge University Press. https://doi.org/10.1017/cbo9780511733017

[14] Swain, M. (1995) Three Functions of Output in Second Language Learning. In: Cook, G. and Seidlhofer, B., Eds., *Principle and Practice in Applied Linguistics*, Oxford University Press, 125-144.

[15] Long, M.H. (1996) The Role of the Linguistic Environment in Second Language Acquisition. In: Ritchie, W.C. and Bhatia, T.K., Eds., *Handbook of Second Language Acquisition*, Elsevier, 413-468. https://doi.org/10.1016/b978-012589042-7/50015-3

[16] Krashen, S.D. (1982) Principles and Practice in Second Language Acquisition. Pergamon.

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

[18] Lantolf, J.P. and Thorne, S.L. (2006) Sociocultural Theory and the Genesis of Second Language Development. Oxford University Press.

[19] Warschauer, M. and Healey, D. (1998) Computers and Language Learning: An Overview. *Language Teaching*, **31**, 57-71. https://doi.org/10.1017/s0261444800012970

[20] Bibauw, S., François, T. and Desmet, P. (2019) Discussing with a Computer to Practice a Foreign Language: Research Synthesis and Conceptual Framework of Dialogue-Based Call. *Computer Assisted Language Learning*, **32**, 827-877. https://doi.org/10.1080/09588221.2018.1535508

[21] McCrocklin, S.M. (2016) Pronunciation Learner Autonomy: The Potential of Automatic Speech Recognition. *System*, **57**, 25-42. https://doi.org/10.1016/j.system.2015.12.013

[22] Godwin-Jones, R. (2021) Big Data and Language Learning: Opportunities and Challenges. *Language Learning & Technology*, **25**, 4-19. https://doi.org/10.64152/10125/44747

[23] Derakhshan, A. and Ghiasvand, F. (2024) Is ChatGPT an Evil or an Angel for Second Language Education and Research? A Phenomenographic Study of Research-Active EFL Teachers' Perceptions. *International Journal of Applied Linguistics*, **34**, 1246-1264. https://doi.org/10.1111/ijal.12561

[24] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., *et al*. (2021) The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*, **372**, n71. https://doi.org/10.1136/bmj.n71

[25] Lee, S., Choe, H., Zou, D. and Jeon, J. (2026) Generative AI (GenAI) in the Language Classroom: A Systematic Review. *Interactive Learning Environments*, **34**, 335-359. https://doi.org/10.1080/10494820.2025.2498537

[26] Liu, J., Sihes, A.J.B. and Lu, Y. (2025) How Do Generative Artificial Intelligence (AI) Tools and Large Language Models (LLMS) Influence Language Learners' Critical Thinking in EFL Education? A Systematic Review. *Smart Learning Environments*, **12**, Article No. 48. https://doi.org/10.1186/s40561-025-00406-0

[27] Wang, C., Li, X., Jin, H. and Zou, B. (2026) Vocal and Visual Design Features in Multimodal Genai Agents: Effects on EFL Speaking Performance, Affect and Interaction Experience. *Computer Assisted Language Learning*, 1-30. https://doi.org/10.1080/09588221.2026.2680421

[28] Behforouz, B., Al Maqbali, A., Al Ghaithi, A. and Poorghorban, A. (2025) Mobile Interaction Meets AI Tutoring: Using ChatGPT-4O to Boost Speaking Skills in EFL Classrooms. *International Journal of Interactive Mobile Technologies* (*iJIM*), **19**, 34-49. https://doi.org/10.3991/ijim.v19i22.57447

[29] Hsu, M., Chen, P. and Yu, C. (2023) Proposing a Task-Oriented Chatbot System for EFL Learners Speaking Practice. *Interactive Learning Environments*, **31**, 4297-4308. https://doi.org/10.1080/10494820.2021.1960864

[30] Yan, H., Singh, M.K.S. and Rawian, R.M. (2024) AI-Based Corrective Feedback in EFL Interactive Speaking: In-Sights from Interactionist SLA Theory. *Educational Sciences*: *Theory & Practice*, **24**, 306-322.

[31] Zou, B., Guan, X., Shao, Y. and Chen, P. (2023) Supporting Speaking Practice by Social Network-Based Interaction in Artificial Intelligence (AI)-Assisted Language Learning. *Sustainability*, **15**, Article 2872. https://doi.org/10.3390/su15042872

[32] Tai, T. (2026) ChatGPT-Assisted EFL Learning Beyond the Classroom: Impact on Oral Proficiency, Enjoyment, and Grit. *Computer Assisted Language Learning*, 1-28. https://doi.org/10.1080/09588221.2026.2640087

[33] Yang, H., Kim, H., Lee, J.H. and Shin, D. (2022) Implementation of an AI Chatbot as an English Conversation Partner in EFL Speaking Classes. *ReCALL*, **34**, 327-343. https://doi.org/10.1017/s0958344022000039

[34] Sayed, B.T., Bani Younes, Z.B., Alkhayyat, A., Adhamova, I. and Teferi, H. (2024) To Be with Artificial Intelligence in Oral Test or Not to Be: A Probe into the Traces of Success in Speaking Skill, Psychological Well-Being, Autonomy, and Academic Buoyancy. *Language Testing in Asia*, **14**, Article No. 49. https://doi.org/10.1186/s40468-024-00321-0

[35] Dandu, G., Charyulu, G.M. and Kumari, K.L. (2024) AI-Driven Language Learning: The Impact of Rosetta Stone on ESL Students' Speaking Proficiency and Self-Control. *Journal on Interdisciplinary Studies in Humanities*, **16**, 1-12.

[36] 吴玲娟. 大语言模型赋能英语口语教学的实证研究[J]. 现代教育技术, 2025, 35(10): 33-41.

[37] 吴坚豪, 周婉婷, 曹超. 生成式人工智能技术赋能口语教学的实证研究[J]. 中国电化教育, 2024(4): 105-111.

[38] Wang, J., Ismail, H.H. and Mansor, A.Z. (2025) Enhancing Chinese EFL Learners' Speaking Proficiency through AI-Integrated POA. *Educational Sciences*: *Theory and Practice*, **25**, 44-57.

[39] Shi, W. and Shakibaei, G. (2025) Insights into the Effectiveness of Artificial Intelligence-Integrated Speaking Instruction in Enhancing Speaking Skills and Social-Emotional Competence as Well as Reducing Demotivation and Shyness. *European Journal of Education*, **60**, e70174. https://doi.org/10.1111/ejed.70174

[40] Çakmak, F. (2022) Chatbot-Human Interaction and Its Effects on EFL Students' L2 Speaking Performance and Anxiety. *Novitas-ROYAL*, **16**, 113-131.

[41] Zheng, Y., Zhou, Y., Chen, X. and Ye, X. (2025) The Influence of Large Language Models as Collaborative Dialogue Partners on EFL English Oral Proficiency and Foreign Language Anxiety. *Computer Assisted Language Learning*, **39**, 1092-1118. https://doi.org/10.1080/09588221.2025.2453191

[42] Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y. and Xing, P. (2024) Exploring EFL Learners' Perceived Promise and Limitations of Using an Artificial Intelligence Speech Evaluation System for Speaking Practice. *System*, **126**, Article 103497. https://doi.org/10.1016/j.system.2024.103497

[43] Lee, J. (2025) Speaking English with AI or Humans: What Engages EFL Learners More? *English Teaching*, **80**, 41-66. https://doi.org/10.15858/engtea.80.5.202512.41

[44] Li, L., Zhang, X., Zou, B. and Yang, Q. (2025) AI Partner or Peer Partner? Exploring AI-Mediated Interaction in EFL Pronunciation from a Socio-Cultural Perspective. *Learning, Culture and Social Interaction*, **55**, Article 100958. https://doi.org/10.1016/j.lcsi.2025.100958

[45] Pituxcoosuvarn, M., Tanimura, M., Murakami, Y. and White, J.S. (2025) Enhancing EFL Speaking Skills with AI-Powered Word Guessing: A Comparison of Human and AI Partners. *Information*, **16**, Article 427. https://doi.org/10.3390/info16060427

[46] Ye, Y., Deng, J., Liang, Q. and Liu, X. (2022) Using a Smartphone-Based Chatbot in EFL Learners' Oral Tasks. *International Journal of Mobile and Blended Learning*, **14**, 1-17. https://doi.org/10.4018/ijmbl.299405

[47] Ericsson, E. and Johansson, S. (2023) English Speaking Practice with Conversational AI: Lower Secondary Students' Educational Experiences over Time. *Computers and Education*: *Artificial Intelligence*, **5**, Article 100164. https://doi.org/10.1016/j.caeai.2023.100164

[48] Lee, J.H., Shin, D. and Hwang, Y. (2024) Investigating the Capabilities of Large Language Model-Based Task-Oriented Dialogue Chatbots from a Learner's Perspective. *System*, **127**, Article 103538.

[49] Shafiee Rad, H. (2024) Revolutionizing L2 Speaking Proficiency, Willingness to Communicate, and Perceptions through Artificial Intelligence: A Case of Speeko Application. *Innovation in Language Learning and Teaching*, **18**, 364-379. https://doi.org/10.1080/17501229.2024.2309539