

在第二外语学习中大语言模型应用的系统文献综述

A Systematic Literature Review of Large Language Models in EFL

魏宏^{1*}, 廖龙燕¹

¹温州大学教师教育学院

* 3563984610@qq.com

【摘要】 随着技术发展推动教育变革, 大语言模型在语言学习领域展现出新活力。研究发现, ChatGPT 等大语言模型在教育中的应用已较为广泛, 但集中在高等教育中, 在 K-12 教育领域的研究仍然有限。本文探讨了 ChatGPT 等大语言模型在 K-12 教育中的实施现状和潜在作用。通过系统文献综述法, 本研究基于 PRISMA 框架, 对 2022 年至 2025 年的相关实证研究进行了文献筛选和分析, 旨在探讨大语言模型支持下的第二外语学习的 AI 角色类型、教育领域、教育阶段、教师角色, 为未来研究提供参考, 助力第二外语教育发展。

【关键词】 第二外语学习; 大语言模型; K-12 教育; 系统文献综述

Abstract: As tech advances reshape education, large language models (LLMs) are becoming increasingly popular in language learning. Research shows LLMs like ChatGPT are widely used in education, mainly in higher education, with limited K-12 application. This paper examines their implementation and potential impact in K-12 education. Using a PRISMA - based systematic literature review, this study analyzed 2022-2025 empirical studies. It focuses on AI roles in LLM - supported second foreign language learning, covering educational domains, stages, and teacher roles. The findings offer insights for future research and practical suggestions for LLM integration in this field, advancing it and providing a reference for educators and researchers.

Keywords: EFL, Large language models, K-12 education, Systematic literature review

1. 引言

随着技术的飞速发展, 人工智能 (AI) 技术正深刻地改变着教育领域。大语言模型 (Large Language Models, LLMs) 作为人工智能的重要分支, 凭借其强大的自然语言处理能力, 为教育带来了新的机遇和挑战。在人工智能研究领域, OpenAI 的 GPT 系列模型已处于前沿地位, 其通过处理海量文本数据所学知识, 为众多任务提供了先进解决方案, 教育领域便是其潜在的重要应用场景之一 (Heidt, 2023)。近年来, LLMs 如 ChatGPT 等在教育中的应用逐渐增多, 其在个性化学习、智能辅导、自动化作业批改等方面展现出巨大潜力, 显著提升了教学效率和学习体验 (卢宇, 余京蕾, 陈鹏鹤, & 李沐云, 2023)。有学者提出, ChatGPT 将会对教育教学理念、教学模式、教育科学研究范式、人才培养等方面带来巨大冲击 (夏立新, 2023)。其优势在于能为学生的论文、作业提供支持, 促进个性化学习, 同时辅助教师的教育教学, 生成教学反馈与评价, 助力教学创新 (AlAfnan, et al, 2023)。然而, 它也存在一些潜在风险, 例如回答问题的准确性不足、伦理与安全问题、知识剽窃和学术不端行为等 (王天恩, 2023)。与此同时, 公众对 ChatGPT+教育的态度是可以提供教育资源, 有助于教育发展 (刘天丽, 杨现民, 李康康, & 糜泽稼, 2023)。这些研究大多证明 ChatGPT 在教育以及学习方面会产生积极影响。本文筛选文献, 旨在关注以 ChatGPT 为首的大语言模型在不同语言技能、不同学段学生中的实施现状。回顾文献综述 (Jaeho Jeon, 2023), 该综述通

过文献计量学分析语音识别聊天机器人在语言学习领域的研究趋势等，发现重点集中在高等教育应用上，指出了 K-12 教育领域的研究缺口。因此，有必要在 K-12 教育领域对第二外语学习中使用以 ChatGPT 为主的大语言模型的研究现状进行全面审查。

以 ChatGPT 为首的基于大语言模型搭建而成的聊天机器人在 EFL(English as a foreign language learning, 英语作为二外的学习)中的研究已经渐渐形成体系。例如，已有研究探讨了语言学习者如何应对与聊天机器人交流时出现的障碍 (Moussalli & Cardoso, 2020)，与聊天机器人的互动如何影响他们的学习成果 (Dizon, 2020)，以及使用聊天机器人对诸如交流意愿、动机和焦虑等情感变量的影响 (Chen, et al, 2020; Jeon, 2022; Tai & Chen, 2020)。总体而言，这些研究证实了使用具备语音识别功能的聊天机器人对语言学习的益处。此外，随着各种由大型语言模型驱动的、具有语音识别功能的聊天机器人在社会中日益普及，它们可能会吸引越来越多的学术界的关注。随着大语言模型与学生和 EFL 之间进行协同，我们需要理解它在协作中的具体角色，以及协作的情况如何演变，为今后更好地使用工具提供指导。于是本文提出研究问题，旨在揭示大语言模型在 EFL 中的角色及协作情况。

根据系统评价和元分析(PRISMA)框架的首选报告项目，对一系列研究进行了检查，以回答以下研究问题：

- RQ1: 在 K-12 教育中，大语言模型支持下的第二外语学习的协作有哪些类型（即大语言模型在协作中的角色）？协作的情况如何？
- RQ2: 在 K-12 教育中，大语言模型应用在哪个教育阶段和哪个语言技能，是如何实施的？实施的结果如何？
- RQ3: 在 K-12 教育中，教师在大语言模型应用到语言学习的角色是什么样的？态度如何？对未来发展有什么建议？

通过这一系统综述，本文期望为大语言模型在 K-12 教育的 EFL 中的应用提供全面的了解，并为未来的研究和实践提供有价值的指导。

2. 文献综述

近年来，随着人工智能和机器学习技术的迅猛发展，像 OpenAI 的 GPT 系列这样的大型语言模型已成为该领域的重要组成部分。这些模型在训练过程中能够“学习”到大量的信息，从而可以执行更复杂的任务，例如回答问题、写作以及编程等。它们通常运用深度学习技术，针对海量的文本数据进行训练，旨在生成与训练数据相似的新文本，或者理解和回答有关训练数据的问题 (吴河江 & 吴砥, 2024)。凭借其多功能性，大型语言模型目前在不同领域都发挥着一定的作用，尤其在第二外语教育领域。

在教育领域的最新趋势中，大型语言模型正逐渐成为各教育层次的重要工具，尤其在低年级阶段，比如小学，其应用已变得更加普遍和关键。在小学五年级中，配备 AI 技术的聊天机器人“Sister Fish”被设计为能够理解 157 本书的基本内容，模仿人类同伴的互动方式，与小学生进行交流以提高学生阅读兴趣 (Liu, 2022)。在小学六年级的英语教育中，教师将大型语言模型 (LLMs) 整合到外语写作教学中，对学生的写作表现、自我调节学习策略使用以及写作动机产生了积极影响 (Liu, 2024)。

随着大语言模型在教育领域的深入应用，其影响力正从传统学科向新兴教育领域快速扩展，不仅赋能传统教学，还为创新教育模式注入了独特价值。这种技术的广泛应用不仅推动了教育手段的革新，也为学生的学习体验和教学效果带来了深远的积极影响。一方面，生成式人工智能 (GAI) 和自动语音识别 (ASR) 技术为外语学习带来了新的工具，尤其是 GAI 聊天机器人。比如用 GAI 聊天机器人 CoolE Bot 对与小学英语学习者进行口语交流。这些技术能够模拟真实的语言环境，提供个性化的学习体验，并给予实时反馈，推动了教育手段的革新 (Tai, et al, 2024)。另一方面，ChatGPT 作为一种非目标导向的聊天机器人，能够显

著支持 EFL 学习者的语言学习过程，尤其是在写作、阅读、语法和词汇方面；也能给学生带来非同寻常且新奇的学习体验，以提高教学效果（Nguyen Hoang, et al, 2024）。

在英语作为第二外语的学习过程中，学习者需要掌握与外语相关的知识和技能，而聊天机器人已被视为一种重要的语言学习工具（Lin & Mubarak, 2021; Yang et al, 2022）。然而，由于课堂时间和资源的限制，教师需要为所有学习者提供及时且有针对性的反馈是一项具有挑战性的任务（Luo, 2016; Wang et al., 2022）。对于外语非母语学习者来说，聊天机器人提供了便捷的学习机会（Jeon, 2024），使他们可以不受时间或地域的限制参与学习（Chen et al., 2020）。学习者可以在模拟场景中与聊天机器人交谈，获得高效、多样化和个性化的学习体验，且不受同伴压力的影响（Moussalli & Cardoso, 2020）。

总之，大语言模型（尤其是 OpenAI 的 GPT 系列）正在教育领域释放出前所未有的潜力。这些模型通过深度学习技术处理海量文本数据，能够精准适配不同教育层次和语言技能的需求。在低年级教育中，它们被广泛应用于提高学生阅读能力的练习活动，并显著提升学生的外语写作能力。与此同时，在口语等语言技能的教学设计中，大语言模型展现了卓越的优化能力，为教学方法注入了创新活力。此外，这些模型不仅极大地改善了学生的学习体验，还通过个性化指导显著提升了学生的批判性思维能力。对于教师而言，它们提供了强大的教学支持工具，帮助教师更高效地设计课程、评估学生表现，并实现因材施教。总的来说，大语言模型在教育领域的应用正释放出巨大的潜力，展现出无限广阔的应用前景。随着技术的不断进步，它们有望成为未来教育生态系统中不可或缺的核心驱动力。

3. 研究方法

系统文献综述法最初定义为“对医疗保健干预效果做出的明确的、可重复的和最新的总结”（Pollock & Berge, 2018），主要应用于医学领域，后逐渐广泛应用于管理研究中。为此，国际研究者制定了 PRISMA 声明（Preferred Reporting Items for Systematic Reviews and Meta-Analyses），以规范系统综述文章的标准。本文使用系统文献综述法，对 2022 至 2025 年关于大语言模型在 K-12 教育 EFL 中应用的文献进行统计分析，并使用 VOSviewer 软件对文献数据进行可视化分析，使研究结果更加直观。

4. 研究过程

在确定研究问题后，本研究遵循以下步骤进行：选择关键词、检索关键词、初步筛选文献、深入分析文献、最终得出研究结果。在文献检索过程中，我们严格遵循 PRISMA（系统综述和荟萃分析的首选报告项目）程序。2021 年国际知名专家团队联合发布了 PRISMA 2020 版声明并更新了条目清单，进一步规范了系统文献综述法的研究流程。初步检索后，我们首先通过审查文献的标题和摘要，筛选出可能符合纳入标准的研究。接下来，对这些初步筛选的文献进行更深入的全文审查，以确定它们是否完全符合系统综述的纳入标准。符合标准的文献将被选中进行详细分析。在整个过程中，研究人员需要准确记录并报告每个步骤中检索、筛选、评估和纳入的研究数量，以及排除研究的具体原因。最后，研究人员将针对之前提出的研究问题进行深入讨论和总结。

4.1. 检索关键词

在制定检索策略时，输入尽可能多的关键术语至关重要，这能避免忽略重要文章。本研究采用布尔搜索策略确定和检索关键词，使用“AND”和“OR”运算符有效检索文献，检索前先对关键词分类，再按类别扩展更多相关关键词，以确保全面检索到所需文章（Wright, Brand, Dunn, & Spindler, 2007）。根据研究的主题，我们将关键词分为“EFL”、“LLMs”和“K-12 education”三大类。在“EFL”这一类中，我们增添了“AI-assisted language learning”、“Language learning”、“Second language”、“Foreign language”、“EFL”、“ESL”、“Language

acquisition”和“Language teaching”这几个关键词，较为全面地涵盖了第二外语学习和教育。在“ChatGPT”这一类中，我们选定了“ChatGPT”、“AIGC”、“Generative AI”、“artificial intelligence”、“OpenAi”、“Large language models”、“LLM”、“LLMs”和“chatbot”，包含了与 ChatGPT 有关的各种概念和相关描述。在“K-12 education”这一类中，我们选取了“K12 education”、“K-12”、“primary education”、“middle school”、“high school”、“kindergarten”，囊括了 K-12 教育的所有学段。在每个类别中，关键词之间用“OR”进行组合，两个类别的关键词之间用“AND”连接。本研究在 Web of science 核心合集数据库中进行检索，考虑到 OpenAi 是在 2022 年开发 ChatGPT，我们新增了“出版时间”这一限制条件，使论文的时间在 2022~2025 年之间，进一步保证论文检索的严谨性。论文检索的条件限制如下图所示。

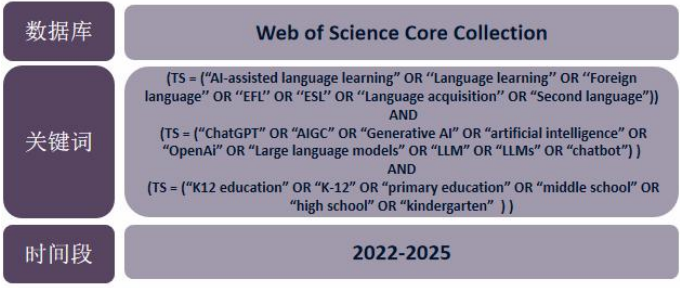


图 1 论文检索条件

4.2. 筛选文献

确定检索范围后，共有 625 篇文献入选，接下来需要将入选的文献进行筛选。在初步筛选中，研究人员排除了 420 篇文献，剩下 205 篇文献需要进一步排查。在二次筛选的过程中，研究人员排除了未明确提及大语言模型、EFL 和 K-12 教育的文献 (n=181)。最后，研究人员对剩下的 24 篇文章进行了第三次筛查，通过精度文献与仔细评估，继续排除了 3 篇论文，其中 2 篇无法检索，1 篇未发表。通过严格的筛查流程（见图 2），最终 21 篇文献被纳入系统综述的数据库中，为后面的数据分析奠定基础，选取和排除的标准如表 1。

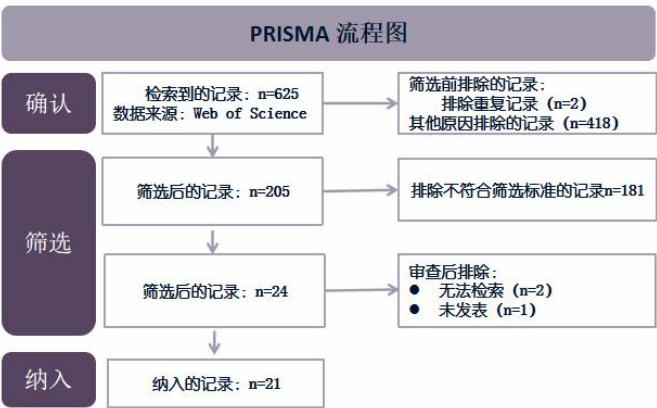


图 2 PRISMA 流程图

表 1 选取和排除标准

选取标准	排除标准
文献内容与 K-12 中的学段相关	文献内容不包含 K-12 任何学段 (即幼儿园、小学、初中、高中)
文献内容有明确提到大语言模型	文献内容并未提及大语言模型
文献内容包含 EFL	文献内容不包含 EFL
文献发表时间于 2022 年及以后	文献发表时间在 2022 年以前

4.3. 分析文献

确定好文献的数据库后，研究人员使用 VOSviewer 进行关键词的聚类分析。VOSviewer 是由荷兰莱顿大学开发的一款文献计量分析软件，也是一个知识图谱可视化工具。它支持标签视图、密度视图、聚类视图和分散视图四种浏览方式，帮助研究者更直观的进行文献分析。文献分析的过程按照“导入数据——清洗数据——生成可视化图谱”的步骤进行。其中在清洗数据的环节，研究人员需要剔除重复表达的数据以保证知识图谱的简洁性与准确性。在生成可视化图谱环节，研究人员需要将同类的关键词整理在同一个聚类，作为研究结果总结的前期准备工作。

5. 研究结果

5.1. 文献基本信息分析

关于文献的分布，符合标准的文献呈现出多样化的地理分布。主要集中在亚洲、欧洲和美洲。这些地区包括英国、美国、中国、韩国、新加坡等。这种分布显示了研究的国际化特征，表明相应的研究话题在全球范围内受到重视。特别值得注意的是，最后筛选出的 21 篇文章全部是实证研究，尤其是那些运用实验法的研究。然而，相比之下，使用访谈法进行的实证研究在入选文献中相对缺乏。

关于文献的研究方法，大部分使用了混合的研究方法，除了做定量的数据分析之外，还针对某些研究问题进行了定性分析。这种方法的优势在于，它能够结合定量数据的客观性和定性数据的深度见解。单独使用定量或定性方法的文献数量也相当，研究者能够在统计验证和深入理解之间取得平衡，从而获得更全面的研究结果。

关于文献的内容，涵盖协作类型、教学水平、语言技能和教师角色四个主要维度。在协作类型方面，ChatGPT 主要作为助教（66.7%，n=14）和学习伙伴（33.3%，n=7）参与教学活动，且作为助教的研究占比显著高于作为学习伙伴的研究。从教学水平来看，文献中讨论的学习者的教育水平涵盖了高中、初中、小学。初等教育（n=7，33.3%）比中等教育（n=12，57.1%）所占的比例略少，除此之外，部分研究所代表的教育水平为初等教育和中等教育混合，统一概括为 K-12，表明中等教育阶段是研究的重点领域。在语言技能分布上，写作技能占比最高（38.1%，n=8），其次是口语（28.5%，n=6）和综合技能（23.8%，n=5），词汇和阅读技能仅各占 4.8%（n=1），显示出写作技能的研究显著多于其他语言技能。至于教师角色，其分布呈现多样化，教师辅助角色占比最高（52.4%，n=11），其次是无教师参与（23.8%，n=5）和教师主导（23.8%，n=5），表明教师在大语言模型相关教学中的辅助性作用受到较多关注。总体而言，研究在协作类型、教学水平、语言技能和教师角色方面均呈现出一定的集中性和不平衡性，为后续研究提供了重要的参考方向。如下表所示。

表 2 文献基本信息表

分类	详细内容	样本量 (%)	对应文献（见附录）
协作类型	作为助教	14 (66.7%)	ID1、ID5、ID6、ID8、ID11、ID12、ID13、ID14、ID15、ID16、ID18、ID19、ID20、ID21
	作为学习伙伴	7 (33.3%)	ID2、ID3、ID4、ID7、ID9、ID10、ID17、

教学水平	初等教育	7 (33.3%)	ID1、ID8、ID9、ID11、ID15、ID16、ID17、
	中等教育	12 (57.1%)	ID4、ID5、ID6、ID7、ID10、ID12、ID13、ID14、ID18、ID19、ID20、ID21
语言技能	K-12 教育	2 (9.6%)	ID2、ID3、
	综合技能	5 (23.8%)	ID2、ID12、ID13、ID15、ID20
	口语	6 (28.5%)	ID3、ID4、ID9、ID16、ID17、ID19、
	写作	8 (38.1%)	ID1、ID5、ID6、ID7、ID8、ID10、ID18、ID21
	词汇	1 (4.8)	ID11
教师角色	阅读	1 (4.8%)	ID14
	教师主导	5 (23.8%)	ID1、ID8、ID11、ID13、ID17
	教师辅助	11 (52.4%)	ID2、ID3、ID4、ID5、ID6、ID9、ID10、ID14、ID15、ID18、ID21
	无教师参与	5 (23.8%)	ID7、ID12、ID16、ID19、ID20

5.2. 文献关键词分析

该关键词的聚类分析图谱系统整合了大语言模型与 K-12 阶段 EFL 教育研究的核心议题，通过技术工具、认知理论、行为机制及教学实践四维框架，揭示了人工智能时代语言教育生态的复杂关联。技术工具维度中，“automated writing evaluation”与“artificial intelligence chatbot corrective feedback”等关键词突显以 ChatGPT 为代表的生成式 AI 在 EFL 教学中的功能定位，其自动化评估与即时反馈机制已实现从辅助工具到教学流程深度整合的范式转变。值得注意的是，“ELIZA”与当代模型的共现不仅映射对话系统技术的历时性演进，更暗示教育技术需在继承性创新中实现场景适配。认知理论维度通过“chinese dual code theory”与“language learning information technology”的关联，阐释了双编码理论的多模态认知机制如何为增强现实语言教学提供科学基础——该理论主张语言符号与非语言表征的协同加工，恰与“augmented reality english”的具身学习设计形成理论-技术闭环。行为机制维度中，“motivation”、“acceptance”等心理变量与“determinants”、“agents”等系统要素的交织，揭示技术落地本质上是多主体参与的复杂适应过程。研究表明，AI 工具效能不仅受算法准确性的单维度制约，更需考量教师技术采纳意愿、学生认知负荷阈值以及课堂技术生态的协同效应。教学效果维度的“accuracy”与“interference”构成辩证张力：大语言模型虽能通过海量语料训练提升语言形式准确性，但其过度纠错可能抑制学习者的容错阈值，而母语负迁移与技术依赖更可能消解“creative writing”所需的元语言意识，这种矛盾凸显 AI 工具需在精准性与教育性之间建立动态平衡。

值得注意的是，图谱的网络化特征在“applications in subject areas language classroom experiences”节点得到集中体现，该枢纽通过连接技术实践、理论框架与教学情境，构建起“技术具身化-理论操作化-场景适应性”的三元结构。这提示教育技术研发需超越简单的工具移植，转而追求教学场景驱动的技术再语境化（recontextualization）过程。

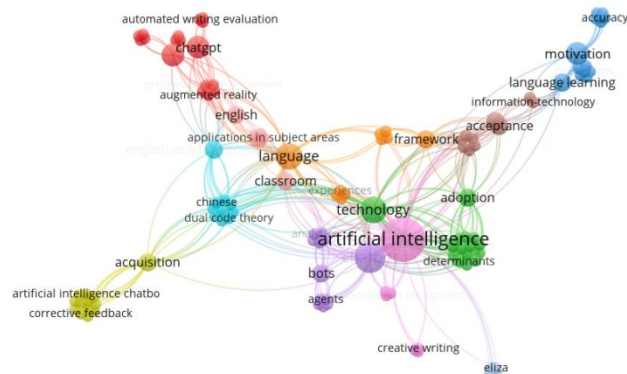


图 3 关键词聚类图谱

密度可视化图谱通过色阶梯度映射了大语言模型在 K-12 阶段 EFL 教育研究的领域聚焦度，其中“automated writing evaluation”、“ChatGPT”及“augmented reality”构成了深色核心区，表明自动化写作评估、生成式 AI 及增强现实技术已成为当前研究的热点。围绕核心区的“corrective feedback”（纠错反馈）、“dual code theory”（双编码理论）与“agents”（智能代理）等次级概念，揭示了技术应用与认知理论深度融合的特征——例如，双编码理论为 AR 英语教学的多模态交互提供了认知框架，而智能代理系统则支撑了个性化反馈的实践落地。图谱的扩展趋势体现在“language classroom experiences”、“information-technology”的色域过渡上，暗示研究重心正从工具效能验证转向技术生态构建，如信息技术如何重塑语言课堂的师生互动范式。值得注意的是，“analytical intelligence chatbot”（分析型智能对话系统）与“determinants”（技术采纳动因）作为新兴节点，其浅色分布反映了前沿探索方向：前者指向超越基础对话功能的认知分析能力开发，后者则关注教师技术接受度的多变量影响机制。而边缘区域的“Eliza”（早期对话系统）与“creative writing”（创意写作）形成鲜明对比，前者作为技术演化的历史锚点，后者则暴露出现有研究对 AI 工具可能抑制语言创造力的隐忧，此类话题的稀疏分布恰凸显了技术伦理研究的不足。

在应用场景维度，核心技术在“subject areas”（学科整合）与“classroom”（课堂教学）中已展现结构化渗透，但在“acquisition”（语言习得机制）与“motivation”（学习动机维持）等深层教育过程的研究密度相对较低。特别是跨文化语境下的技术适配问题（如“chinese”节点的孤立性），以及 AI 反馈对元语言意识发展的长周期影响，尚未形成系统化研究网络。未来亟需在“human-AI collaborative framework”（人机协同框架）构建、“cognitive load optimization”（认知负荷优化）及“ethical boundary design”（伦理边界设计）等方向深化探索，这些空白领域或将决定大语言模型能否真正实现从“教学工具”到“教育范式”的质变跨越。

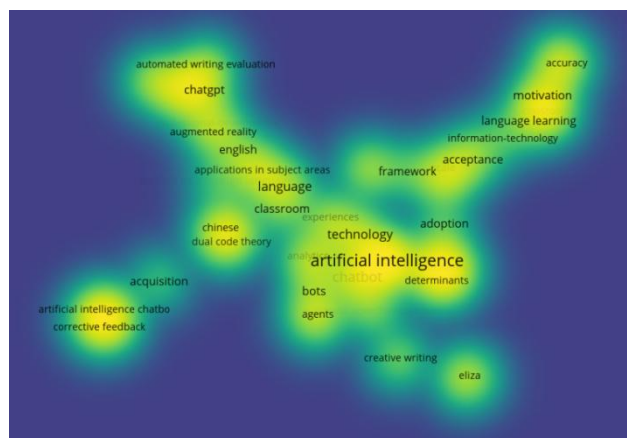


图 4 密度可视化图谱

6. 讨论

6.1. RQ1: 在 K-12 教育中，大语言模型支持下的第二外语学习的协作有哪些类型（即大语言模型在协作中的角色）？协作的情况如何？

在 K-12 教育中，大语言模型（LLMs）的协作类型主要分为两类，其中以助教角色为主以学习伙伴为辅。一方面，大语言模型在协作过程中以助教身份提供支持。例如，在写作训练中，大语言模型作为辅助工具，嵌入认知学术语言学习方法（CALLA, Cognitive Academic Language Learning Approach）框架，通过个性化支持、实时反馈和适应性交互，针对学习者的错误提供纠正性重铸反馈，帮助学习者练习和掌握英语致使移动结构，从而辅助语言学习并支持教师完成教学任务。然而，这种助教角色并未完全替代教师的主导地位，也不同于平等交流的学习伙伴角色，而是以辅助身份协助学生进行写作练习（Liu et al, 2024）。此外，在词汇学习中，基于人工智能的物体检测翻译（AI-based ODT）应用程序以助教身份为学生提供词汇学习支持，包括识别物体、翻译和提供发音等，但并不主导教学过程，而是辅助学生完成词汇学习任务（Liu & Chen, 2023）。在阅读练习中，聊天机器人（ChatGPT）通过提示工程技术（prompt engineering）协助教师自动生成英语阅读理解问题，减轻教师负担，以助教身份支持教学活动（Lee et al, 2024）。此外，“AI 教练”（AI Coach）作为一种虚拟代理，通过个性化反馈帮助学生进行口语和听力练习、学习词汇以及纠正发音问题，从而培养学生的语言综合技能（Wang et al, 2023）。另一方面，大语言模型以学习伙伴的身份陪伴学生学习。它能够与学生互动交流，提供学习资源、协助完成学习任务，并满足日常交流需求，从而支持学生的自主学习。例如，学生在使用 ChatGPT 时，将其视为能够提供帮助和支持的伙伴（Nguyen et al, 2024）。此外，聊天机器人“Ellie”被设计为与学生进行英语对话的伙伴，通过鼓励学生参与交流，辅助提升其语言能力，承担学习伙伴的角色，以提高学生的口语交际能力（Yang et al, 2022）。

总之，大语言模型作为一种强大的人机协作工具，在以上两个方面发挥着积极的作用。作为助教，可以提供个性化支持、实时反馈和适应性交互，帮助学生掌握语言结构，辅助教师完成教学任务，但不替代教师的主导地位；作为学习伙伴，通过互动交流和资源支持，帮助学生完成学习任务，促进自主学习。这种双重协作模式表明，大语言模型在 K-12 教育中既可以作为助教提供教学支持，也可以作为学习伙伴促进学生的自主学习。

这些研究发现揭示了大语言模型在 K-12 教育中的 EFL 作为协作工具的多样化应用潜力。其多重协作类型为教育创新提供了多维视角，推动了个性化学习的发展，并促进了学生的全面成长。综合来看，大语言模型在 K-12 教育中的 EFL 应用不仅丰富了教学实践，还预示着教育领域的持续创新与进步。

6.2. RQ2: 在 K-12 教育中, 大语言模型应用在哪个教育阶段和哪个语言技能, 是如何实施的? 实施的结果如何?

大语言模型在 EFL 领域应用多样化, 涉及不同教育阶段以及不同语言技能, 包括: 在高中生口语培训方面, 聊天机器人“Ellie”通过提供对话机会和任务导向的交流来支持韩国高中生的口语技能提升 (Kim et al, 2022); 在高中写作领域, 基于 ChatGPT 的聊天机器人 (KR-Automated Essay Scoring System) 主要用于辅助教师对学生的英语作文进行评分和提供反馈, 通过分析学生作文的任务完成情况、内容、结构和语言使用等方面, 给与学生及时、个性化的反馈 (Shin1 & Lee, 2024; Luo et al, 2025); 在小学口语训练中, Dialogflow 开发的聊天机器人与学生进行结构化、重复性对话, 提供反馈, 帮助学生练习英语口语, 在练习过程中如同伙伴一样与学生互动交流, 通过互动式口语测试等方式评估学生的口语能力, 重点关注口语技能的提升 (Jeon & Lee, 2024)。

研究主要集中在中等教育阶段和写作领域, 而初等教育甚至幼儿园教育的应用则相对较少, 尤其是在词汇和阅读能力培养方面, 仍存在较大的空白。这些发现不仅展示了大语言模型在教育领域的应用潜力, 也揭示了其在特定场景中的局限性。尽管研究覆盖了中小学教育阶段, 但幼儿园阶段的应用似乎较为欠缺。这可能是因为幼儿园教育更注重亲身体验和人际交往, 对机器学习工具的需求相对较低。然而, 大语言模型的应用并不仅限于学生, 还涉及教育者。大语言模型在辅助教育者设计课程和开发评估工具方面具有显著潜力 (Cooper, 2023)。

在实施过程中, 大语言模型在提升学生技能、情感和学业表现方面展现出显著的积极作用。具体而言, 在技能方面, 基于人工智能的英语学习支持系统通过分析文本和视频内容生成个性化学习材料, 帮助学生高效学习英语词汇; 同时, 系统提供的学习计划和自主学习功能促使学生深入阅读和理解英语文本, 而测试模式及内容整理分享活动则进一步锻炼了学生的写作与表达能力, 最终全面提升其语言综合应用能力 (Lee et al, 2023)。情感层面, 人工智能交互支架 (AIIS) 凭借动态反馈与个性化支持, 助力学生设定学习目标并激发学习兴趣, 从而显著增强其内在学习动力。这种动机提升与积极情感体验紧密相连, 表明 AIIS 能够有效改善学生在非正式学习环境中的情感状态 (Wang et al, 2025)。在学业表现上, 聊天机器人写作练习对学生的写作能力提升效果显著。实验数据显示, 实验组学生的后测写作成绩明显优于对照组, 充分证明了聊天机器人在写作教学中的应用价值。此外, 实验组学生普遍反馈, 聊天机器人有助于复习课堂内容、提高写作技能, 且在使用过程中减少了对犯错的焦虑感, 表现出更高的学习舒适度 (Kwon et al, 2023)。总之, 大语言模型可以通过个性化学习支持、情感激励和互动练习, 显著提升学生技能、情感和学业表现, 助力教育创新与效率提升。

6.3. RQ3: 在 K-12 教育中, 教师在大语言模型应用到语言学习的角色是什么样的? 态度如何? 对未来发展有什么建议?

教师在大语言模型应用到语言学习的角色是主要是辅助角色, 还有部分是教师主导或者无教师参与。比如教师为学生介绍 ChatGPT、教授提示工程技能、提供指导和翻译等, 但学生是在教师引导下自主与聊天机器人互动完成写作任务, 给与学生学习自主权的同时减轻教师负担 (Woo et al, 2024; Tai & Chen, 2024)。同时, 人工智能 (AI) 在英语教学中的应用, AI 的角色是作为助教, 教师是作为教学的主导者。教师在教学过程中决定是否使用 AI 以及如何使用 AI 来辅助教学或者教授学生零样本提示方法和数字写作技能等, 引导学生利用工具写作。更有甚者无教师参与, 比如人工智能辅助语言学习 (AILL) 为学生提供如自动评分、反馈、个性化指导等帮助, 辅助学生进行英语学习, 再如 PengTalk 的语音识别聊天机器人帮助学生练习发音、纠正错误, 辅助学生进行英语学习, 但并不主导教学过程。且无教师参与, 由此让学生自主学习、实现真正的教育放权 (An et al, 2023)。

大语言模型对 K-12 教育中的 EFL 的影响是复杂多样的。研究表明, 教师对 ChatGPT 的

认知程度和使用态度仍存在差异。部分教师对 AI 工具持欢迎态度,认为其能减轻教学管理和写作批改的工作量;但也有教师持谨慎态度,对 AI 工具的能力有所怀疑(Luo.et al,2025)。此外,教师们普遍认为 ChatGPT 在生成问题时表现出一定的有效性,但也存在一些问题。例如,ChatGPT 生成的“是/否”问题因句子结构复杂,可能分散学生注意力,导致学生花费更多时间理解问题结构而非内容。因此,教师建议在使用 ChatGPT 生成问题时提供更详细的提示和指导,并对生成的问题进行验证和调整,特别是在复杂任务和推断性问题上,以提高问题的准确性和相关性(Lee.et al,2024)。还有的建议教师要了解学生向生成式 AI 工具提问的目的,从而提供有针对性的指导和支持,帮助学生更好地利用 AI 工具进行写作(Woo.et al,2024)。

总体而言,教师和学生大语言模型在 EFL 中的应用持复杂态度。对于其辅助教学设计、提升学习体验和学习动机的潜力,师生普遍持积极看法,并希望通过进一步的学习与培训,更好地将其作为一种学习工具加以利用。然而,在信息准确性等问题上,师生则表现出审慎甚至负面的态度,认为大语言模型可能对学习环境构成潜在风险。因此,大语言模型在教育中的应用亟需进一步规范,未来的研究应填补其在实证研究中的空白,以形成对大语言模型在 EFL 中更全面、更深入的认识。

7. 总结

本研究通过对 2022 年至 2025 年间发表的同行评议文章进行系统回顾和元分析,深入探讨了大语言模型在 K-12 教育阶段英语作为外语教学中的应用效果。研究结果显示,大语言模型在显著提升学习者英语学习成效方面发挥了重要作用,不仅有效促进了学习者的语言技能发展,还在提升其心理和情感幸福感方面展现出积极影响。这些发现为大语言模型在 EFL 教育中的应用提供了有力的实证支持,表明其在优化教学方法和学习体验方面具有巨大潜力。

本次研究揭示了大语言模型在 K-12 教育中 EFL 应用的不足,主要体现在对幼儿教育等阶段及某些语言技能研究的缺失,以及对技术潜在风险关注的不足。未来研究需聚焦于技术在这些薄弱领域的应用,并重视教育公平和学术诚信。同时,技术开发应围绕教育核心价值,如依学习理论开发多模态学习工具,利用增强现实技术打造沉浸式学习环境。还需构建全面的技术伦理框架,平衡效率与人文关怀,如设计反馈敏感度调节器,通过边缘计算缩小数字鸿沟。此外,跨学科研究方法创新至关重要,建议采用设计性研究范式,在真实教育场景中迭代验证 AI 工具,并构建教师-AI 协同教学模型,重新定义教师角色。总之,只有技术逻辑与教育逻辑深度融合,人工智能才能真正推动 EFL 教育革新。

参考文献

- 夏立新.ChatGPT 对教育的多重变[J].国家教育行政学院学报,2023,(3):9-12.
- 刘天丽,杨现民,李康康,& 糜泽稼.(2023). 社会公众如何看待“ChatGPT+教育”?——基于国内主流开放社区的文本分析.现代教育技术,33(10),14-23.
- 卢宇,余京蕾,陈鹏鹤,& 李沐云.(2023). 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例.中国远程教育,43(4),24-31,51.
- 王天恩.ChatGPT 的特性、教育意义及其问题应对[J].思想理论教育,2023,(4):19-25.
- 吴河江,吴砥.(2024). 生成式人工智能教育应用: 发展历史、国际态势与未来展望.比较教育研究,2024(6),13-23.
- AlAfnan M A, Dishari S, Jovic M, et al. ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses[J]. Journal of Artificial Intelligence and Technology, 2023,(2):60-68.
- An, X., Chai, C. S., Li, Y., Zhou, Y., Shen, X., Zheng, C., & Chen, M. (2023). Modeling

English teachers' behavioral intention to use artificial intelligence in middle schools. *Education and Information Technologies*, 28, 5187 - 5208.

An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*.

Chen, H. H. J., Yang, C. T. Y., & Lai, K. K. W. (2020). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*. Advance online publication.

Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A ChatBot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, 58(6), 1161–1189.

Cooper, G. (2023b). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, 32(3), 444–452.

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology*, 24(1), 16–26. <https://doi.org/10125/44705>

Heidt, A. (2023). 'Arms race with automation': professors fret about AI-generated coursework. *Nature*.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873–15892.

Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*. Advance online publication.

Jeon, J., Lee, S., & Coronel-Molina, S. M. (2024). Rethinking AI: bias in speech-recognition chatbots for ELT. *ELT Journal*, 78(4), 435 - 445.

Jeon, J., & Lee, S. (2024). The impact of a chatbot-assisted flipped approach on EFL learner interaction. *Educational Technology & Society*, 27(4), 218-234.

Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). Design principles and architecture of a second language learning chatbot. *Language Learning & Technology*, 26(1), 1–18.

Kim, R. (2024). Effects of learner uptake following automatic corrective recast from Artificial Intelligence chatbots on the learning of English caused motion construction. *Language Learning & Technology*, 28(2), 109–133.

Kwon, S. K., Shin, D., & Lee, Y. (2023). The application of chatbot as an L2 writing practice tool. *Language Learning & Technology*, 27(1), 1 - 19.

Lee, D., & Kim, H. (2023). Development research on an AI English learning support system to facilitate learner-generated context-based learning. *Education Technology & Development*, 71(6), 629–666.

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, 29, 11483 - 11515.

Lin, C. J., & Mubarak, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, 24(4), 16–35.

Liu, C. C., Liao, M. G., Chang, C. H., & Lin, H. M. (2022). An analysis of children's interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, 189, 104576.

Liu, P.-L., & Chen, C.-J. (2023). Using an AI-Based Object Detection Translation Application for English Vocabulary Learning. *Educational Technology & Society*, 26(3), 5-20.

Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2024). Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 1-25.

Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for

efficient classroom instruction. *Computer Assisted Language Learning*, 29(3), 451-476.

Luo, M., Hu, X., & Zhong, C. (2025). The collaboration of AI and teacher in feedback provision and its impact on EFL learners' argumentative writing. *Education and Information Technologies*.

Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, 33(8), 865–890.

Pollock, A., & Berge, E. (2018). How to do a systematic review. *International Journal of Stroke*, 13(2), 138–156.

Shin, D., & Lee, J. H. (2024). Exploratory study on the potential of ChatGPT as a rater of second language writing. *Education and Information Technologies*, 29, 24735–24757.

Tai, T. Y., & Chen, H. H. J. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*. Advance online publication.

Tai, T. Y., & Chen, H. H. J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, 220, 105112.

Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System*, 103528.

Wang, F., Zhou, X., Li, K., Cheung, A. C. K., & Tian, M. (2025). The effects of artificial intelligence-based interactive scaffolding on secondary students' speaking performance, goal setting, self-evaluation, and motivation in informal digital learning of English. *Interactive Learning Environments*.

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, 194, 104703.

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, 37(4), 814–840.

Woo, D. J., Susanto, H., Yeung, C. H., Guo, K., & Fung, A. K. Y. (2024). Exploring AI-Generated text in student writing: How does AI help? *Language Learning & Technology*, 28(2), 183–209.

Woo, D. J., Wang, D., Susanto, H., & Wallace, Y. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction. *Education and Information Technologies*, 29, 24963 – 24990.

Wright, R. W., Brand, R. A., Dunn, W., & Spindler, K. P. (2007). How to Write a Systematic Review: *Clinical Orthopaedics and Related Research*, 455, 23–29.

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343.

附录

ID	出版年份	研究方法	文献名
1	2024	定量研究法	Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation
2	2024	混合研究法	ChatGPT as a tool for self-learning English

			among EFL learners: A multi-methods study
3	2022	定量研究法	Implementation of an AI chatbot as an English conversation partner in EFL speaking classes
4	2022	定量研究法	Design principles and architecture of a second language learning chatbot
5	2024	定量研究法	Effects of learner uptake following automatic corrective recast from Artificial Intelligence chatbots on the learning of English caused-motion construction
6	2024	定量研究法	Exploratory study on the potential of ChatGPT as a rater of second language writing
7	2024	定性研究法	Exploring EFL students' prompt engineering in human-AI story writing: an activity theory perspective
8	2023	混合研究法	The application of chatbot as an L2 writing practice tool
9	2024	混合研究法	Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions
10	2024	混合研究方法	Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process
11	2023	定量研究法	Using an AI-Based Object Detection Translation Application for English Vocabulary Learning
12	2023	定量研究法	Modeling students' perceptions of artificial intelligence assisted language learning
13	2023		

14	2024	定量研究方	Modeling English teachers' behavioral intention to use artificial intelligence in middle schools
15	2023	混合研究法	Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english education
16	2024	混合研究法	What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis
17	2024	混合研究法	Rethinking AI: bias in speechrecognition chatbots for ELT
18	2024	混合研究法	The impact of a chatbot-assisted flipped approach on EFL learner interaction
19	2025	混合研究法	Exploring AI-Generated text in student writing: How does AI help?
20	2023	混合研究法	The effects of artificial intelligence-based interactive scaffolding on secondary students' speaking performance, goal setting, self-evaluation, and motivation in informal digital learning of English
21	2025	混合研究法	Development research on an AI English learning support system to facilitate learner-generated-context-based learning
		定量研究法	The collaboration of AI and teacher in feedback provision and its impact on EFL learner's argumentative writing
