

什么样的聊天机器人更能促进学习?

——基于60项实验与准实验研究的元分析

马玉慧^{1,2} 杨海珍² 刘凯^{2,3}

(1. 曲阜师范大学传媒学院, 山东日照 276800; 2. 渤海大学教育科学学院, 辽宁锦州 121013;
3. 渤海大学通用人工智能研究所, 辽宁锦州 121013)

[摘要] 随着人工智能技术的迅猛发展,特别是大语言模型的兴起,应用聊天机器人辅助学生学习成为研究热点。相关研究表明,聊天机器人能有效激发学生的学习动机,提高学习投入度、自我效能感和学业成就等。然而,究竟什么样的聊天机器人更能促进学习,如何才能有效地应用聊天机器人,仍是亟待解决的问题。本研究采用元分析方法,系统分析近十年国内外应用聊天机器人促进学生学习的60项实证研究。结果表明,聊天机器人对学生学习效果有中等程度的正向影响,反馈类型、交互模式等交互设计对学习效果有显著的调节作用。不同年级、不同学习方式对聊天机器人的应用效果存在显著差异,但不同学科和学习环境无显著差异。文章最后展望了聊天机器人的未来研究方向。

[关键词] 聊天机器人; 大语言模型; 学习效果; 元分析

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2024)06-0079-09

一、引言

对话是人类社会交往中最常用、最直接的交互方式。聊天机器人能使用自然语言与人类对话,在一定情境下可代替人类完成某些任务。因此,应用聊天机器人模拟教师与学生进行沟通交流,始终是教育研究者追求的目标(屈静等, 2020)。早在1966年,研究者就设计和开发了聊天机器人ELIZA(戴岭等, 2023),用来模拟心理治疗师,通过询问问题与对话,对患者进行心理治疗。这一创举开启了聊天机器人的相关研究。但之后的几十年,

聊天机器人在教育领域并未得到广泛应用,主要原因在于技术条件的限制与较高的开发成本。

近年来,随着人工智能技术的飞速发展,特别是大语言模型的出现,聊天机器人在用户对话意图的识别和对话连贯性、准确性等方面都有极大改进,用户体验大幅提升(吴砥等, 2023)。聊天机器人再度成为教育研究者关注的焦点(Labadze et al., 2023; 王琦等, 2024)。聊天机器人能够模拟教师、同伴等,通过文本、语音等多种交互方式与学生对话,为学生提供个性化学习内容与反馈,满足个性化学习需求(Sikström et al., 2022)。目前聊天机器

[收稿日期] 2024-09-20

[修回日期] 2024-11-06

[DOI编码] 10.13966/j.cnki.kfjyyj.2024.06.010

[基金项目] 教育部人文社会科学研究一般项目“聊天机器人对话策略及其对学生数学学习的影响研究”(23YJAZH099)。

[作者简介] 马玉慧, 博士, 教授, 硕士生导师, 曲阜师范大学传媒学院, 研究方向: 人工智能教育应用(799493385@qq.com); 杨海珍, 研究生, 渤海大学教育科学学院, 研究方向: 人工智能教育应用; 刘凯(通讯作者), 博士, 副教授, 硕士生导师, 渤海大学通用人工智能研究所, 研究方向: 通用人工智能、机器教育、精神病理学(ccnulk@ccnu.edu.cn)。

[引用信息] 马玉慧, 杨海珍, 刘凯(2024). 什么样的聊天机器人更能促进学习? ——基于60项实验与准实验研究的元分析[J]. 开放教育研究, 30(6): 79-87.

人已被应用于多个学科领域。

研究者围绕聊天机器人对学习效果的影响展开了探讨,但结论并不一致(Dai et al., 2022)。为此,有研究者利用元分析探讨了聊天机器人对学习效果的影响(Alemdag, 2023; Wu et al., 2024; Deng, 2023),但对什么样的聊天机器人,以及如何应用聊天机器人更能促进学生学习并未深入分析。此外,已有元分析研究(Wu et al., 2024)未能涵盖大语言模型方面的研究成果。本研究采用元分析方法,对已有聊天机器人教育应用的实证研究进行梳理,重点探讨聊天机器人的人机对话方式和应用模式对学生学习效果的影响。

二、相关研究

(一)潜在调节变量

1. 交互模态

当前,学生与聊天机器人的交互模态已从单一的文本交互发展到语音和多模态交互。文本是人机交互最基础的交互模态(Rapp et al., 2021)。它简单、直接,不受时空限制,学习者可随时随地输入文本与聊天机器人对话。语音输入是近年发展起来的交互模态,学生可通过语音与聊天机器人对话。语音交互更接近日常教学的对话,即使是儿童也可通过语音与聊天机器人交互,从而节省时间(Xu et al., 2021; Jeon et al., 2023a; Jeon et al., 2023b)。多模态不仅可以提高交互的丰富性和趣味性,还能通过多感官刺激,提升学习者的学习体验(Aslan et al., 2024; Tai et al., 2024)。但目前聊天机器人中不同输入模态对学习效果影响的研究结论并不一致。温克勒等(Winkler et al., 2020)探究了使用文本输入和文本+语音输入对学生学习编程的影响。结果表明,使用文本+语音输入比单一的文本输入更有利于学生学习。利特曼等(Litman et al., 2006)的研究结果则表明,使用聊天机器人辅导学生学习,输入模态由文本换成语音对学习效果的影响并无显著差异。因此,不同交互模态对学习效果的影响有待进一步探究。

2. 主导类型

根据对话发起者的不同,聊天机器人可分为聊天机器人主导和学生主导两类(Kuhail et al., 2023)。聊天机器人主导指聊天机器人是对话的发起者和

主导者。交互过程中,聊天机器人根据特定学习任务采用向学生提出问题,引导、启发学生思考或干预学习过程。学生对聊天机器人提出的问题进行响应和回答,聊天机器人则在给予反馈后,开始新一轮会话(Ruan et al., 2020; Winkler et al., 2020)。这类聊天机器人通常由设计者预先设定启发性问题和反馈。学生主导的聊天机器人指对话由学生发起,学生主动向聊天机器人发起对话或提出问题,聊天机器人进行响应,学生再根据聊天机器人的回复继续对话或开始新一轮对话(Han et al., 2022)。常见的问题回答类聊天机器人和基于大语言模型的聊天机器人都属这一类。

比较而言,两类聊天机器人各有优势。聊天机器人主导的对话可以基于某种特定教学理论提供学习支架(Ruan et al., 2020; Winkler et al., 2020),精心设计对话过程,特别是设计提问与反馈。这种对话方式类似课堂上由教师主导的师生互动。交互过程中,聊天机器人把控对话走向,通过提出引导性问题不断引领学生逼近学习目标,完成学习任务。而学生主导的对话中,学生掌握主动权,聊天机器人对学生的提问给予反馈,更能有效地解决个性化问题。目前学界尚不清楚两类聊天机器人在学习效果上是否有显著差异。

3. 反馈类型

反馈是学习过程的关键环节。无论在何种学习环境,反馈都能影响学习效果。大量研究结果表明,不同的反馈类型对学习效果的影响存在差异(Cai et al., 2023)。舒特(Shute, 2008)根据反馈内容,将反馈分为简单反馈和详细反馈。简单反馈指仅提供学习结果的对错,不包含额外的解释或指导。简单反馈又可分为“结果反馈”(knowledge of results, KR)和“纠错反馈”(knowledge of correct response, KCR)。结果反馈指仅告知学生回答正确或错误,是最简单的反馈方式。纠错反馈指不仅告知正误,还提供正确答案。详细反馈指提供详尽的反馈信息,包括解释性反馈(explanation feedback, ExF)、提示反馈(prompt feedback, PF)和元认知反馈(metacognitive feedback, MF)三类。解释性反馈指为学生提供答案并做进一步解释,提示反馈指通过提供例子、提示或问题等启发学生思考,明确下一步该如何解决问题,元认知反馈从元认知层面,

通过提示或提出问题引发学生反思和监管学习过程。已有元分析研究探索了技术丰富型学习环境中不同反馈类型对学习效果的影响(Cai et al., 2023)。研究表明,详细反馈比简单反馈有效。在三类详细反馈中,解释性反馈最有效。尽管大量研究分析了不同学习情境下反馈对学习的影响,但尚不清楚聊天机器人中不同反馈类型对学习效果的影响。

4. 学生年级

鉴于不同年级学生认知特点不同,已有聊天机器人的元分析研究探讨了学生年级对聊天机器人应用效果的影响,但结论不一致。阿勒姆达格(Alemdag, 2023)认为 K-12 学生和大学生使用聊天机器人学习的效果并无显著差异。然而,有学者(Wu et al., 2024)通过元分析研究发现,相对于小学生和中学生,聊天机器人对大学生的影响更大。因此,基于更多的实证研究结果,将年级作为一个调节变量,探究对聊天机器人应用对学习效果是否有显著影响尤为重要。

5. 学科类别

目前,聊天机器人已广泛应用于多个学科(Hwang et al., 2023)。库海勒等(Kuhail et al., 2023)对 2011—2021 年间的相关文献开展了统计分析,发现聊天机器人主要应用在计算机科学、语言学习和数学等学科,其中计算机科学的应用研究最多,超过 30%。元分析结果表明,聊天机器人在不同学科中的应用效果并无显著差异(Deng & Yu, 2023; Alemdag, 2023)。

有研究者(Li et al., 2022)将学科类别分为理论知识和实践技能两类。理论知识课程侧重定义、定理等知识的学习,通常采用测验或考试的方式评价学习结果。实践技能课程侧重在完成项目或各项学习任务过程中培养技能,学习结果的评价常基于项目或任务完成过程中的表现或作品。本研究重点探究聊天机器人对理论知识学习和实践技能学习的效果是否存在显著差异。

6. 学习环境

将聊天机器人应用于不同学习环境,会对学习产生不同的影响(Winkler et al., 2018)。目前,聊天机器人的应用环境可分为两类:混合学习和在线学习。混合学习中,聊天机器人被视为课堂学习的一

部分,学生应用聊天机器人完成教师布置的学习任务,与其他教学活动融为一体(Ait Baha et al., 2024)。随着聊天机器人应用的推广,课堂上应用聊天机器人逐渐增多,如郭等(Guo et al., 2023)应用聊天机器人帮助学生提高辩论技能。学生应用聊天机器人进行练习,构建支持观点的证据,并预测可能的反驳意见。研究表明,辩论前应用聊天机器人可以使学生的辩论更有序、充分和详实。在线学习环境中学生可在聊天机器人辅助下进行在线学习,无需教师面对面参与(Winkler et al., 2020),如在 MOOC 平台嵌入聊天机器人,帮助学生解答学习常见问题,提高学习体验。结果表明,学生应用聊天机器人解答常见问题的体验不如使用网页方式(Han et al., 2022)。也有研究(Chen et al., 2020)发现,一对一环境下聊天机器人对学习的影响显著高于教室环境。因此,目前学习环境对聊天机器人应用效果的影响有无显著差异尚不清楚。

7. 学习方式

应用聊天机器人学习的方式有两种:自主学习和协作学习。自主学习指学生使用聊天机器人进行个性化学习。学生按照自己的步调学习,自己选定学习内容,或使用聊天机器人个性化推荐的学习资源。学生与聊天机器人之间通过一对一对话完成学习任务。这是目前聊天机器人的主要应用方式(Huang et al., 2022)。聊天机器人还可以作为协作小组成员,扮演知识贫乏的同伴或教师参与小组学习,引导学生思考,确保协作过程顺利完成(Nguyen, 2023)。已有研究表明,个人和小组的学习效果都有提高(Tegos & Demetriadis, 2017)。但目前尚未有研究关注自主学习和协作学习中应用聊天机器人对学习效果的影响有无显著差异。

(二)相关元分析研究

已有研究者对聊天机器人的学习效果及影响因素进行了元分析研究。例如,邓和于(Deng & Yu, 2023)采用元分析方法分析 2010—2022 年间发表的 32 项实证研究发现,聊天机器人在促进学习效果方面展现出正向的积极影响,效应值达 0.79。张等(Zhang et al., 2023)探讨聊天机器人在辅助语言学习方面的有效性发现,平均效应处于中等水平,效应值为 0.527。吴和于(Wu & Yu, 2024)的元分析结果表明,AI 聊天机器人对学生学习效果有显著

影响。这些发现证实了聊天机器人在教育领域的巨大应用潜力和价值。

尽管已有元分析研究探讨了聊天机器人的应用效果和影响因素,但就聊天机器人的发展与应用态势而言,仍有必要深入探究。首先,已有元分析数据截至 2022 年,研究数据量较小,且多集中在高等教育,K-12 的应用较少(Alemdag, 2023; Huang et al., 2022)。其次,2022 年后,聊天机器人可视化开发平台得到推广使用,再加上大语言模型的发布,其在教育领域的应用研究迅速增多。因此,在更多研究的基础上对聊天机器人开展更完整、更综合的元分析非常必要。再次,已有聊天机器人元分析研究未将交互设计、应用方式等因素纳入分析范畴(Alemdag, 2023, Wu et al., 2024)。这些因素对设计与应用聊天机器人都具有重要影响。为此,本研究采用元分析方法,对近十年相关文献进行分析,着重解决两个问题:

1)聊天机器人中什么样的交互设计,包括交互模态、反馈方式、主导类型更能促进学习?

2)聊天机器人的何种应用方式,包括学习环境、学习方式、学生学段、学科类别更能促进学习?

三、研究设计

(一)研究过程

本研究以“chatbot”“conversational agent”“conversational tutor”“ChatGPT”“gpt”“generated artificial intelligence”“google bard”“generated AI”“聊天机器人”“智能导学系统”“智能学习系统”“教育机器人”“大模型”“生成式人工智能”为主题和关键词,检索 Elsevier、Web of Science、Taylor & Francis、Wiley、Springer、Google Scholar 和中国知网等数据库,检索时间为 2013 年 1 月 1 日至 2024 年 6 月 23 日,共获得 28486 项研究。

(二)文献纳入标准

聊天机器人文献的纳入标准包括:1)研究主题为聊天机器人对学生学习的影响;2)采用实验或准实验研究方法;3)实验组使用聊天机器人,对照组未使用;4)研究报告包含可用于计算效应量的值,如实验组与对照组的样本量、均值、标准差或 t 值。本次元分析排除会议记录、新闻、文献综述、非聊

天机器人对学习的影响、非实证研究、不能获取全文、研究数据不完整以及实验组和对照组都使用聊天机器人等类型的文献,最终纳入 60 篇文献。检索与筛选过程见图 1。

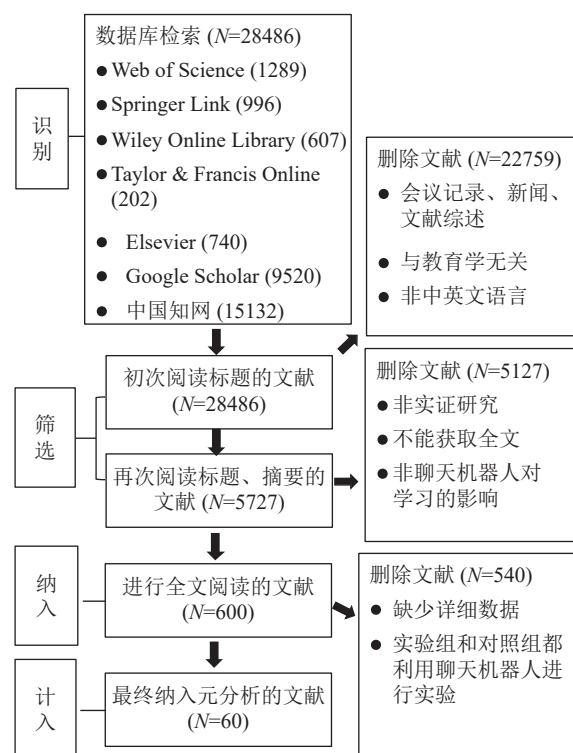


图1 文献检索与筛选过程

(三)文献编码

文献编码包括文献作者、年份、实验人数、学生学段(学前、K-12、大学)、学科类别(理论知识、实践技能)、交互模态(文本、文本+语音、语音、多模态)、主导类型(聊天机器人主导、学生主导)、学习环境(混合学习、在线学习)、学习方式(一对一自主学习、协作学习)和反馈类型(简单反馈、详细反馈、混合反馈)。其中,简单反馈和详细反馈参照舒特(Shute, 2008)的分类。混合反馈包括两种及以上的详细反馈,如解释性反馈+提示反馈、解释性反馈+元认知反馈等。

(四)数据分析

本研究选用 Comprehensive Meta Analysis 3.0 (CMA3.0)软件作为元分析工具,鉴于样本量和研究数量相对较小,采用 Hedges's g(简称 g 值)作为核心的最终效应值,在 60 项确定的研究文献中计算了 80 个效应值。元分析的两个核心模型架构为

固定效应模型与随机效应模型。考虑到各文献的特征变量显著影响元分析结果的稳健性与普适性,本研究采用随机效应模型,确保分析结论的科学严谨性与适用性。

四、元分析结果

(一)发表偏倚检验

本研究使用漏斗图(见图 2)、线性回归检验和秩相关检验失安全系数(见表 1)进行发表偏差检验。漏斗图上的点基本围绕合并效应值 0.436 对称散开。Begg 检验的 p 值为 0.088, Egger's 检验 p 值为 $0.06 > 0.05$, 失安全系数 Nfs 值为 4932, 超过检验标准 $5 \times 80 + 10$ 。由此可见,本研究不存在发表偏倚现象,样本有效可靠。

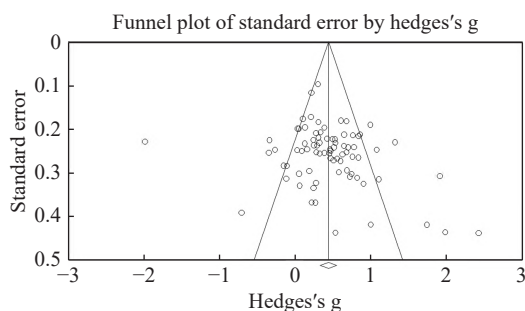


图 2 学习效果漏斗图

表 1 发表偏倚检验

研究数	Rosenthal's Nfs	Begg		Egger's intercept	
		Z	P	T	P
80	4932	1.708	0.088	1.909	0.060

(二)异质性检验

异质性检验采用 I^2 检验和 Q 检验进行分析。检验标准为:当 $0 < I^2 < 40$ 时,存在低程度的异质性;当 $40 < I^2 < 60$ 时,存在中度异质性;当 $60 < I^2 < 75$ 时,存在较大的异质性;当 $75 < I^2 < 100$ 时,存在极大的异质性(Higgins et al., 2003), Q 值的显著性 $p < 0.05$ 。异质性检验结果见表 2。其中, $Q = 326.181$ ($p = 0.000 < 0.001$), $I^2 = 75.780$ 。这说明各样本之间存在异质性,可以采用随机效应模型进行研究。

(三)聊天机器人对学习整体影响

本研究使用 Hedges's g (g 值)衡量聊天机器人对学习效果的影响,根据 Cohen 标准,效应值小于 0.2 为小效应,0.2 至 0.8 为中等效应,大于 0.8 为大

表 2 异质性检验结果

效应模型	研究数	异质性检验			
		Q 值	df	p	I^2
随机效应模型(REM)	80	326.181	79	***	75.780

注:***表示 $p < 0.001$ 。

效应(见表 3)。聊天机器人的效应值为 0.436, $p < 0.05$,表明其对学生学习有中等程度的正向影响。

表 3 元分析合并效应值

效应模型	研究数	合并效应值	95% 置信区间		双尾检验	
			下限	上限	Z	P
随机效应模型(REM)	80	0.436	0.328	0.543	7.931	***

注:***表示 $p < 0.001$ 。

(四)调节变量效果检验

研究结果表明,不同交互模态对学生学习效果的影响存在显著差异。文本与语音结合的交互模态(效应量 $g = 0.881$, $p < 0.05$)显示出较强的正向影响,其效果显著优于纯文本交互模态(效应量 $g = 0.635$, $p < 0.05$)和多模态交互模态(效应量 $g = 0.305$, $p < 0.05$)。单纯的语音交互模态(效应量 $g = -0.026$, $p = 0.887$)未对学生学习产生显著影响。此外,本研究通过分析组间效应,得到 QBET 值为 25.281,且 p 值为 0.000,这证实不同交互模态在促进学生学习方面存在显著差异。

不同反馈类型对学生学习影响方面,混合反馈($g = 0.636$, $p < 0.05$)>详细反馈($g = 0.327$, $p < 0.05$),但简单反馈($g = 0.173$, $p = 0.522 > 0.05$)的影响不显著。这表明前两种聊天机器人的反馈方式对学生学习均产生了中等程度的正向影响,混合方式对学生学习影响最大,简单反馈没有显著影响。组间效应分析结果为:QBET=10.202 和 $p = 0.006 < 0.05$,这说明聊天机器人反馈类型对学生学习影响有显著差异。

聊天机器人的主导类型对学生学习效果的影响无显著差异。从研究结果可知,学生主导($g = 0.472$, $p < 0.05$)和机器人主导($g = 0.385$, $p < 0.05$)对学习效果均存在中等程度的正向影响。组间异质性检验结果为: $Q = 0.674$, $p = 0.412 > 0.05$,这表明二者无显著差异。

在学生学段维度,聊天机器人对大学生($g = 0.518$, $p < 0.05$)和 K-12 学生($g = 0.278$, $p = 0.044 < 0.05$)

的正向影响显著,但对学前儿童($g=-0.198, p=0.290>0.05$)无显著影响。此外,组间异质性检验结果为: $Q=15.234, p=0.000<0.05$,这说明聊天机器人对不同学段学生学习效果的影响具有显著差异。

在混合学习环境($g=0.445, p<0.05$)和在线学习环境下($g=0.372, p=0.009<0.05$),聊天机器人对学习效果均有显著影响。组间异质性检验结果为: $Q=0.218, p=0.640>0.05$,这说明聊天机器人的应用环境对学习效果的影响无显著差异。

协作学习($g=0.668, p<0.05$)和一对一自主学习($g=0.391, p<0.05$)的学习效果均存在中等程度的正向影响。组间异质性检验结果为: $Q=3.930, p=0.047<0.05$,这表明聊天机器人对协作学习和自主学习的影响存在显著差异。

从学科类别看,聊天机器人对理论知识类课程($g=0.454, p<0.05$)和实践技能类课程($g=0.409, p<0.05$)的学习效果均存在中等程度的正向影响。组间异质性检验结果为: $Q=0.196, p=0.658>0.05$,这表明二者无显著差异。

五、结论与讨论

(一) 结论

本研究采用元分析方法分析2013—2024年间聊天机器人对学生学习影响的相关实验或准实验研究,得出以下结论:1)聊天机器人对学生学习有中等程度的正向影响;2)聊天机器人的反馈类型、交互模态、学习方式,以及学生学段都对学习效果有调节效应,但聊天机器人的主导类型、学科类别和学习环境对学习效果无调节效应。

(二) 讨论

1. 聊天机器人对学习效果的整体现象

主效应检验结果表明,聊天机器人对学习效果有显著的正向影响,总效应值为0.396,达到显著水平。这一发现与部分研究结论(Deng & Yu, 2023; Zhang et al., 2023)相呼应。聊天机器人之所以能有效提升学习效果,原因可归结为:1)学生能使用自然语言,以对话方式与聊天机器人交互。学生或提出问题,或回答聊天机器人提问。对话过程能激发学习者的认知参与,使学习者从被动接受信息转变为积极思考,从而提高对学习材料的理解和记忆(Kuhail et al., 2023; Mageira et al., 2022)。2)聊天机

器人能及时响应学生个性化学习需求,为学生提供个性化支持与帮助,使学习问题与障碍得到解决,因而有助于增强学习动机,缓解学习焦虑与压力(Wu & Yu, 2024)。3)当学生回答正确时,及时反馈能强化学习;若学生回答错误,反馈有助于及时纠正错误。这种反馈即时性使强化作用得到更好的发挥,提高学习动机和效果(Melanko & Larkin, 2013)。本研究证实了聊天机器人对学生学业成绩的促进作用。

2. 调节变量对学习效果的调节影响

在聊天机器人与学生的交互模态中,文本与语音相结合的方式、单纯的文本交互,以及多模态交互方式都能对学生学习产生显著正向影响。文本与语音相结合的效果要优于单纯的文本交互,单纯的文本交互效果优于多模态交互,但语音交互对学习效果无显著影响。根据梅耶的多媒体学习理论,视觉和听觉双通道刺激有利于加深记忆和理解,学习效果比单纯的视觉刺激好(Mayer & Moreno, 2003)。本研究证实了学习的双通道理论。以多模态方式进行交互,虽然为学生提供了多感官刺激,如同时呈现文字、视频或动画、声音,但他们需要对这些信息进行有效整合。整合过程可能会在一定程度上产生认知负荷,分散学生注意力,进而影响学习效果(吴丹等, 2024)。若仅通过语音进行交互,虽然输入更快捷,但缺少视觉支持,学生较难在没有文本支持的情况下完成知识建构。此外,语音信息的转瞬即逝特点可能会使学生难以在没有视觉刺激下保持注意力。

本研究结果表明,聊天机器人的不同反馈方式对学习效果的调节影响有显著性差异。其中,混合反馈方式,即解释性反馈+提示反馈或解释性反馈+元认知反馈的效果最好,其次是解释性反馈,但简单反馈对学习效果无显著影响。这与部分研究结论(Cai et al., 2023)相似。认知负荷理论能解释其原因(Sweller et al., 1998)。简单反馈虽然给出答案,但未对答案作解释,反而可能增加学生工作记忆负担,因此交互过程仅给出简单反馈并不能有效促进学习。解释性反馈的详细解释有助于降低学生的外部无关认知负荷,增加相关认知负荷,因此对学习效果有促进作用。这种方式通常给出完整答案并解释说明,学生需将外部知识整合到自身知识体系,某种程度上会带来认知负荷。混合反馈常根据

教学理论开展反馈设计, 并给出详细解释, 还会呈现提示性问题, 或元认知问题作为学习支架, 这样可以引导学生分解学习任务, 启发学生思考(Ruan et al., 2020; Winkler et al., 2020), 降低学习过程的外部认知负荷。此外, 分解后的学习任务降低了学生处理信息的难度, 降低了内在认知负荷, 学习效果会更好。

交互过程中, 学生主导和聊天机器人主导在学习效果方面并无显著差异, 两种方式都能促进学生学习。比较而言, 学生主导的对话模式效果略好。这与部分研究结论(Zhang et al., 2023)一致。自我决定理论(Self-Determination Theory, SDT)表明, 若个体的自主性、效能感和关联感得以满足, 学生的内在动机更易被激发(Deci & Ryan, 2013)。学生主导的对话允许学生有话语权并能根据个人的理解控制对话节奏。因此, 在人机协作的环境中, 学生倾向于自己主导交互方式(Oh et al., 2018)。但这种对话方式要求学生能清晰描述问题, 对他们提出更高要求。相反, 聊天机器人主导的学习方式按预定的对话路径展开, 对话过程可能僵化, 且预定的内容可能无法满足学生的个性化需求。但这些对话路径往往是由设计者精心安排的。因此, 这种模式可能更适用于不能清晰表达学习需求的学习者。究竟采用哪种主导类型有待深入研究。

不同学段的学生应用聊天机器人的学习效果差异显著。聊天机器人对大学生学习效果的影响最大, 其次是 K-12 学生, 对学前儿童的影响最小。这一研究结论证实了部分学者(Wu & Yu, 2024)的研究结果, 但不支持阿勒姆达格(Alemdag, 2023)的研究结论。大学生在技术支持的学习环境中有更强的自我效能感和自主调节能力(Xia et al., 2023; Garzón et al., 2019)。他们更能清晰表达自己的学习需求或观点, 而 K-12 学生和学前儿童更倾向于在教师引导下学习。

学科类别的影响方面, 聊天机器人无论对理论知识的学习, 还是实践技能的提高均有显著影响, 且对两者的影响无显著差异, 这说明聊天机器人既能促进理论知识的学习, 又能促进实践技能的提高。这一结论与部分研究结论(Deng & Yu, 2023; Alemdag, 2023)一致。

聊天机器人在混合学习和在线学习环境下对

学习效果均有促进作用, 但无显著差异。相对于在线学习, 混合学习环境因有教师参与, 学生的学习投入度可能更高, 注意力更集中, 但应用聊天机器人的时间会受教师教学活动安排的限制。而在在线学习环境中, 学生可以根据自己的学习需求决定聊天机器人的参与时机和时长, 从而更好地实现个性化学习。因此, 在混合学习和在线学习环境中应用聊天机器人各有优势, 应用效果无显著差异。

聊天机器人既可用于一对一的学生自主学习, 也可用于协作学习, 但协作学习的应用效果更好。一对一学习中, 聊天机器人的及时反馈和个性化学习支持, 能更好地满足学生个性化学习需求, 有助于提高学习动机和学习投入度, 减少认知负荷(Sikström et al., 2022)。在协作学习环境中, 聊天机器人通过协作支架, 促进学生间交流, 提升讨论质量, 进而提高个人和小组学习效果(Nguyen, 2023)。相对于自主学习, 协作学习的人际交往有助于促进学生的认知发展, 减少学习焦虑和孤独感(李宝敏, 2019)。协作过程中易出现小组成员参与不均衡、学生合作意识差、小组冲突等问题, 而聊天机器人能监控小组学习进度, 引导协作, 促进参与(Tegos & Demetriadis, 2017)。因此, 聊天机器人参与的协作学习能取得更好的学习效果。

(三) 不足与展望

本研究不足之处在于: 第一, 聊天机器人的设计和应用中影响学习效果的因素很多, 本研究未能探讨学生个体差异, 如学习动机、先验知识水平等对学习效果的影响, 这些变量可能对研究结果产生影响; 第二, 部分调节变量的界定可能存在偏差, 如部分文献未能清晰描述反馈类型, 对这些数据的处理可能存在偏差。

未来研究可从以下方面改进: 首先, 探究聊天机器人对不同学习者的影响, 如个体在认知或情感等方面的差异如何影响学习效果; 其次, 深入探讨针对某具体学科内容、不同设计策略对学习效果的影响; 最后, 需要更多的实证研究来验证聊天机器人在不同教育环境的有效性, 以及学习者数据对聊天机器人训练的反哺价值(刘凯, 2023)。

[参考文献]

- [1] Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2024). The impact of educational chatbot on student learning experience[J]. Edu-

cation and Information Technologies, 29(8): 10153-10176.

[2] Alemdag, E. (2023). The effect of chatbots on learning: A meta-analysis of empirical research[J]. Journal of Research on Technology in Education, 1-23.

[3] Aslan, S., Durham, L. M., Alyuz, N., Chierichetti, R., Denman, P. A., Okur, E., Gonzalez Aguirre, D. I., Zamora Esquivel, J. C., Cordourier Maruri, H. A., Sharma, S., Raffa, G., Mayer, R. E., & Nachman, L. (2024). What is the impact of a multi-modal pedagogical conversational AI system on parents' concerns about technology use by young children?[J]. British Journal of Educational Technology, 55(4): 1625-1650.

[4] Cai, Z., Gui, Y., Mao, P., Wang, Z., Hao, X., Fan, X., & Tai, R. H. (2023). The effect of feedback on academic achievement in technology-rich learning environments (TREs): A meta-analytic review[J]. Educational Research Review, 39: 100521.

[5] Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A chatbot for learning Chinese: Learning achievement and technology acceptance[J]. Journal of Educational Computing Research, 58(6): 1161-1189.

[6] 戴岭, 赵晓伟, 祝智庭(2023). 智慧问学: 基于 ChatGPT 的对话式学习新模式[J]. 开放教育研究, 29(6): 42-51+111.

[7] Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects[J]. Computers & Education, 190: 104607.

[8] Deci, E. L., & Ryan, R. M. (2013). Intrinsic motivation and self-determination in human behavior[M]. New York: Plenum. Springer Science & Business Media; 11-40.

[9] Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education[J]. Sustainability, 15(4): 2940.

[10] Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of Augmented Reality on students' learning gains[J]. Educational Research Review, 27: 244-260.

[11] Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation[J]. Computers & Education, 203: 104862.

[12] Han, S., & Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses[J]. Computers & Education, 179: 104395.

[13] Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses[J]. British Medical Journal, 327(7414): 557-560.

[14] Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning[J]. Journal of Computer Assisted Learning, 38(1): 237-257.

[15] Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education[J]. Interactive Learning Environments, 31(7): 4099-4112.

[16] Jeon, J., Lee, S., & Choe, H. (2023a). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning[J]. Computers & Education, 206: 104898.

[17] Jeon, J., Lee, S., & Choi, S. (2023b). A systematic review of research on speech-recognition chatbots for language learning: Implica-

tions for future directions in the era of large language models[J]. Interactive Learning Environments, 1-19.

[18] Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review[J]. Education and Information Technologies, 28(1): 973-1018.

[19] Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review[J]. International Journal of Educational Technology in Higher Education, 20(1): 56.

[20] Li, L., Johnson, J., Aarhus, W., & Shah, D. (2022). Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit[J]. Computers & Education, 176: 104354.

[21] Litman, D. J., Rosé, C. P., Forbes-Riley, K., VanLehn, K., Bhembé, D., & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring[J]. International Journal of Artificial Intelligence in Education, 16(2): 145-170.

[22] 李宝敏, 宫玲玲(2019). 合作学习对学生学习成效的影响研究——基于国内外 54 项实验研究和准实验研究的元分析[J]. 教育发展研究, 39(24): 39-47.

[23] 刘凯(2023). 人工智能与教育学融合的双重范式变革[J]. 开放教育研究, 29(3): 4-18.

[24] Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning[J]. Applied Sciences, 12(7): 3239.

[25] Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning[J]. Educational psychologist, 38(1): 43-52.

[26] Melanko, S., & Larkin, K. T. (2013). RETRACTED ARTICLE: Preference for immediate reinforcement over delayed reinforcement: Relation between delay discounting and health behavior[J]. Journal of behavioral medicine, 36: 34-43.

[27] Nguyen, H. (2023). Role design considerations of conversational agents to facilitate discussion and systems thinking[J]. Computers & Education, 192: 104661.

[28] Oh, C., Song, J., Choi, J., Kim, S., Lee, S., & Suh, B. (2018). I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence[C]//Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems: 1-13.

[29] 屈静, 刘凯, 胡祥恩, 杨钊, 蒋卓轩(2020). 对话式智能导学系统研究现状及趋势[J]. 开放教育研究, 26(4): 112-120.

[30] Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots[J]. International Journal of Human-Computer Studies, 151: 102630.

[31] Ruan, S., He, J., Ying, R., Burkle, J., Hakim, D., Wang, A., Yin, Y., Zhou, L., Xu, Q., AbuHashem, A., Dietz, G., Murnane, E. L., Brunskill, E., & Landay, J. A. (2020). Supporting children's math learning with feedback-augmented narrative technology[C]//Proceedings of the interaction design and children conference: 567-580.

[32] Shute, V. J. (2008). Focus on formative feedback[J]. Review of educational research, 78(1): 153-189.

[33] Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review[J]. Computers & Education, 188: 104564.

- [34] Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design[J]. *Educational psychology review*, 10; 251-296.
- [35] Tai, T. Y., & Chen, H. H. J. (2024). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension [J]. *Computer Assisted Language Learning*, 37(3): 433-460.
- [36] Tegos, S., & Demetriadis, S. (2017). Conversational agents improve peer learning through building on prior knowledge[J]. *Journal of Educational Technology & Society*, 20(1): 99-111.
- [37] 王琦, 郭芳, 张艺馨, 马思腾(2024). 基于大模型的深度课堂变革研究 [J]. *开放教育研究*, 30(4): 104-112.
- [38] Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent[C]//*Proceedings of the 2020 CHI conference on human factors in computing systems*: 1-14.
- [39] Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis[J]. *Academy of Management Proceedings*, 2018(1): 15903-15943.
- [40] 吴丹, 林蕙(2024). 基于脑电频谱的公共数字文化多模态资源认知负荷研究——以国家公共文化云为例 [J/OL]. *情报科学*, 1-23[2024-10-23]. <http://kns.cnki.net/kcms/detail/22.1264.G2.20240729.1541.009.html>.
- [41] 吴砥, 李环, 陈旭(2023). 人工智能通用大模型教育应用影响探析 [J]. *开放教育研究*, 29(2): 19-25+45.
- [42] Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis[J]. *British Journal of Educational Technology*, 55(1): 10-33.
- [43] Xia, Q., Chiu, T. K., Chai, C. S., & Xie, K. (2023). The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot[J]. *British Journal of Educational Technology*, 54(4): 967-986.
- [44] Xu, Y., Wang, D., & Collins, P. (2021). Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner[J]. *Computers & Education*, 161: 104059.
- [45] Zhang, S., Shan, C., Lee, J. S. Y., Che, S., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis[J]. *Education and Information Technologies*, 28(11): 15223-152.

(编辑: 李学书)

What's the Better Chatbot in Learning? --A Meta-analysis of 60 Experiments and Quasi-experiments Studies

MA Yuhui^{1,2}, YANG Haizhen² & LIU Kai^{2,3}

(1. Department of Communication, Qufu Normal University, Rizhao 276800, China;

2. Department of Educational Science, Bohai University, Jingzhou 121013, China;

3. Institute of Artificial General Intelligence, Bohai University, Jingzhou 121013, China)

Abstract: With the rapid development of artificial intelligence technology with large language models, the application of chatbots to assist student learning has become a research hotspot. Relevant studies have shown that chatbots can effectively stimulate students' learning motivation and improve their engagement, self-efficacy, academic achievement, etc. However, what kind of chatbots can promote learning more effectively and how to apply chatbots more effectively are still urgent problems to be solved. To explore these issues, this study uses the meta-analysis method to systematically analyze the 60 empirical studies on applying chatbots to promote student learning conducted internationally in the past decade. The analysis results show that chatbots have a moderate positive impact on students' learning outcomes. In addition, the interactive design of chatbots, including feedback types and interaction modalities, significantly moderates learning outcomes. At the same time, we also found that the grades of students and the application context of chatbots also significantly affect the application outcomes. However, the dominant mode of interaction, subject discipline, and learning mode have no significant effects on learning outcomes. Finally, the article looks forward to future research on chatbots, hoping to provide more guidance and inspiration for the education field.

Key words: Chatbots; large language modeling; learning effectiveness; meta-analysis