



大语言模型在医学统计软件课程辅助教学中的应用效果评价

王 钰,温绮岚,陈津伟,常义坤,吴跃千,唐 慧,张王剑,杜志成
中山大学公共卫生学院医学统计学系/中山大学全球卫生研究中心/中山大学卫生信息
研究中心,广东 广州 510080

摘要:目的 评价基于大语言模型(Large Language Models, LLM)辅助医学统计软件课程的教学改革效果。方法 开课前采用问卷调查学生学习背景,并在课中进行LLM的介绍与使用教育。结课后利用问卷调查,结合学生的课程成绩,多角度评估改革效果。结果 LLM引入课程后,更多学生将LLM列为答疑解惑的选择($P<0.001$);报告的统计软件学习困难的学生人数下降(引入前85.25% vs. 引入后77.97%);LLM回答准确性得到更多学生的认可(引入前45.90% vs. 引入后64.40%, $P=0.016$)。在LLM使用态度上,学生对LLM提升学习兴趣的态度转变不显著(引入前59.02% vs. 引入后50.85%, $P=0.714$);对LLM帮助学习的态度转变明显(引入前77.05% vs. 引入后54.24%, $P=0.018$)。LLM引入课程后50.85%的学生报告学习主动性提升;67.79%的学生愿意将LLM融入未来学习并推荐给身边学生。结论 改革后的LLM教学融合,在增加学生学习途径的同时,有效促进了风险意识的建立,提升学生思辨能力,预防LLM过度依赖。

关键词:医学统计软件;本科生教育;大语言模型;教学改革

DOI:10.13555/j.cnki.c.m.e.2026.04.010

中图分类号:G642 文献标志码:A 文章编号:2096-3181(2026)04-0585-09

Application Effectiveness Evaluation of Large Language Models in Assisting Teaching for Medical Statistics Software Course

WANG Yu, WEN Qilan, CHEN Jinwei, CHANG Yikun, WU Yueqian, TANG Hui,
ZHANG Wangjian, DU Zhicheng

Department of Medical Statistics, School of Public Health, Sun Yat-sen University/Global Health
Research Center, Sun Yat-sen University/Health Information Research Center, Sun Yat-sen
University, Guangzhou 510080, China

Abstract: *Objective* To evaluate the effectiveness of the teaching reform integrating large language models (LLM) into the medical statistics software course. *Methods* A questionnaire survey was conducted before the course to investigate students' learning backgrounds, and instruction on the introduction and use of LLM

收稿日期:2025-10-22

基金项目:中山大学本科教学质量与教学改革工程类项目(51000-12253008)

作者简介:王 钰(2002—),女,江西吉安人,硕士,主要研究方向为传染病疾病负担。

通信作者:张王剑(1989—),男,浙江台州人,副教授,博士,主要研究方向为医学统计学;

杜志成(1989—),男,广东东莞人,副教授,博士,主要研究方向为传染病监测预警统计建模方法及应用、常用医学统计方法及应用。



was provided during the course. After the course, a post-course questionnaire was administered, and the results were combined with students' final grades for a comprehensive evaluation. *Results* After introducing LLM into the course, more students chose LLM as an option for answering questions and solving problems ($P<0.001$). The proportion of students reporting difficulties in learning statistical software decreased (before: 85.25% vs. after: 77.97%). More students recognized the accuracy of LLM-generated answers (before: 45.90% vs. after: 64.40%, $P=0.016$). Regarding attitudes toward the use of LLM, there was no significant change in students' attitudes toward the role of LLM in enhancing learning interest (before: 59.02% vs. after: 50.85%, $P=0.714$), whereas a significant change was observed in their attitudes toward the role of LLM in facilitating learning (before: 77.05% vs. after: 54.24%, $P=0.018$). After the introduction of LLM into the course, 50.85% of students reported improved learning initiative, and 67.79% were willing to integrate LLM into their future learning and recommend them to their classmates. *Conclusion* The integration of LLM into the reformed curriculum not only expanded students' learning pathways but also effectively promoted the development of risk awareness, enhanced critical thinking skills, and prevented over-reliance on LLM.

Keywords: Medical statistics software; Undergraduate education; Large language models; Teaching reform

医学统计学是一门对数据进行收集、整理、分析、解释和展示的学科^[1],是医学生的基础必修课程,也是医学科研实践的重要工具。统计软件应用则是沟通理论和实践的桥梁,通过SAS、SPSS、R、Python等专业软件,学生将书中统计原理转化为实际分析能力,处理现实世界的具体问题。其中,R语言以其易用性和强大功能备受推崇,在医学专业中采用R语言进行教学,能够帮助学生更好地学习和应用统计分析^[2-3]。然而,传统课堂教育虽然满足大多数学生的需求,但在满足个人学习偏好和风格方面存在局限^[4]。实践教学,课堂时间限制学生吸收理解全部知识;软件实操中的个性化问题难以被捕捉。

大语言模型(Large Language Models, LLM)是基于深度学习的自然语言处理模型,能够通过学习大量的语言数据来理解和生成自然语言文本,对用户输入生成对话式响应^[5]。随着信息技术地飞速发展,数智化时代到来,LLM的技术经历几十年的演变,如今已在日常生活中被广泛使用,活跃在公众视野中。如ChatGPT、Kimi ai、DeepSeek、智谱清言、文小言、讯飞星火、腾讯元宝等生成式人工智能(Generative Artificial Intelligence, GAI)。此外,在数智化浪潮的推动下,课堂教学形态经历着从传统向数智化的深刻转型,课堂教学形态正逐步走向人本化、智能化与个性化^[6]。LLM的易获取性、实时性和个性化反馈,为医学和健康科学的教育工作者提供了新思路^[7]。

目前,国内外有很多学者开始探索将LLM引入课程教学,美国的一项研究发现,LLM可协助准备医

师执照考试的医学生轻松理解问题文本中包含的关键信息、逻辑和信息关系的方向性,拥有部分教授医学知识的能力^[8]。Dai等人的研究发现LLM的句子释义能力和广泛的知识储备可以帮助有阅读障碍的学生学习^[9],充当学习助手,及时回答问题,并用简单的句子解释复杂的术语和做法^[10]。医学教育和培训可以使用LLM作为增强传统课程和教学的工具,这些模型可以为实习医师提供真实的临床场景和反馈,从而加强临床教育^[4,11-12]。此外,在软件编程方面,Shue等发现LLM的生成能力能帮助学生克服编程障碍,提高批判性思维和评估LLM响应的能力^[13]。目前也有许多LLM的编程教学尝试借助模型优化常规的算法教学流程,让教师更专注于教授高级的算法^[14]。

在此背景下,如何将GAI-LLM整合到学习和教学实践中,成功辅助教学成为新议题。然在医学统计软件教学的相关课程中正缺少这样的尝试,目前的研究对教学中LLM的应用和评估有限。本研究基于普适性LLM为基础,构建医学统计专用的大语言模型“医学统计助教”,应用于统计软件教学中,并在课程中实施指导和干预,评估LLM的应用对学生和教学效果的实际影响。旨在培养学生的信息检索和实践反思能力,激发学生学习兴趣,同时解决在教学过程中教师解答实践问题的时间和地点限制,探索LLM辅助教学的有效途径。

1 研究方法

1.1 大语言模型的介绍

LLM使用来自科研文章、书籍和其他基于网



络内容的数十亿个词组进行训练。通常,LLM利用深度学习来表示基于文本的训练数据集中单词之间的复杂关联关系,通过多阶段,不同程度地人工输入,LLM学习词组在语言中如何相互使用,并可以应用这些学习的模式来完成自然语言处理任务^[15-16]。本研究所构建的微调GAI,目标是生成可以按需创建内容并与应用程序中的自然语言处理相交的模型。

我们所期待的GAI在对话过程中要扮演的角色是“苏格拉底导师模式”,它鼓励学生通过以递减的难度提出问题来独立思考,直到学生能够解决手头更完整的问题^[10]。

1.2 大语言模型的构建及在课程中的应用

本研究基于北京智谱华章科技有限公司推出的生成式AI助手“智谱清言”为基础,整合更多医学统计相关的原创知识点资料,构建医学统计专用的大语言模型“医学统计助教”。该模型的构建原理及主要步骤:

①模型通过大量文本数据的训练,学习语言的规律和结构;②用户提出问题,模型解析并理解其意图;

③模型结合已学到的知识和技能,生成相应的回答。我们在此基础上将知识库配置增设团队撰写的200多篇医学统计咨询案例原创知识点资料(约11万字),以及R语言实战教材电子版(约29万字)。

使该模型成为医学统计专用的LLM,尤其适用于医学统计软件R语言的编程答疑。学生可以通过扫描二维码访问小程序的方式使用该模型进行问题咨询,实现全天候医学统计助教的功能(使用链接:<https://chatglm.cn/share/FqZMZ>)。截至2025年11月,累计4 626人使用,7 563次对话,本次教学实践中涉及使用学生75人。

本课程将LLM集成到教学活动和流程中,分为课前介绍、课上指导和课后反思三个阶段,见图1。课程改革聚焦构建“批判性LLM教学融合”方式:实施“透明化”教学,主动揭示LLM的潜力与风险,将风险认知转化为具体防御技能;其次积极引导学生建立“LLM初步解答”到“标记存疑”“权威验证”,最后“讨论辨析”的学习闭环;最后强化“人本”对话与反思,通过小组讨论弥补AI的互动局限,将学习重心从获取答案回归到锤炼思维。

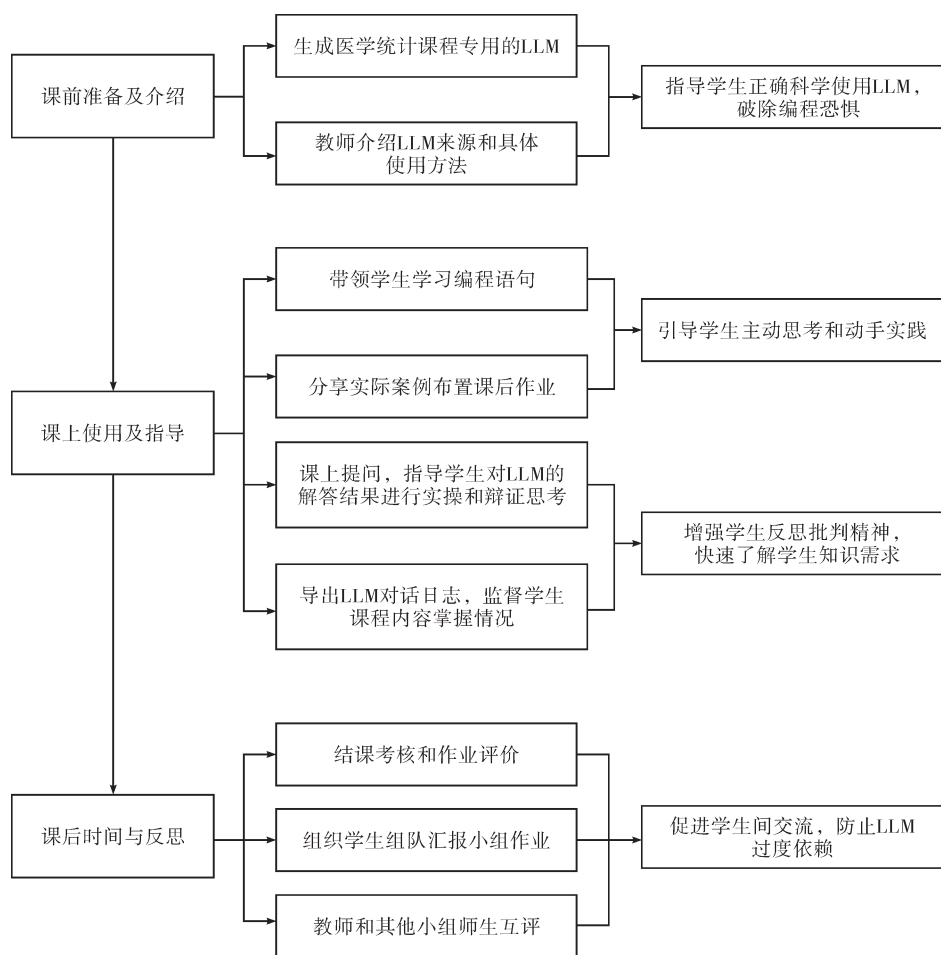


图1 教学流程



1.3 评价方式与统计学方法

采用问卷调查的方式,通过问卷星组卷,在课程前后发放。采用调查问卷综合结课成绩评估本次教学改革效果,进行LLM实际应用效果评估。问卷由本课题组成员设计,基于李克特5级量表展开调查,包含了五个强度等级,比如:非常同意、同意、一般、不同意、非常不同意。

所有统计分析采用R软件(版本4.4.2)完成。计数资料以 $n(\%)$ 表示,将多选题的选项设置为二分类变量,课程前后差异性比较采用 χ^2 检验,当有单元格理论频数 <5 时,采用Fisher确切概率法;有序分类的

选择题采用秩和检验分析差异性, $P<0.05$ 认为差异有统计学意义。

2 教学改革方案实施效果

选取2024—2025学年第一学期,“常用统计软件应用”课程教学改革方案在中山大学的预防医学本科生中实施效果,课程共36学时。

2.1 学生课前情况调查

课程开始前,向学生发放匿名调查问卷,评估学生学习基础、对LLM的了解、日常解决问题的方式和课程期待等情况,见表1。包括16个问题,共61人参与调查,回收率为81.33%,有效率为100%。

表1 选课学生课前情况调查

题目	选项[n(%)]				
1. 接触过哪些医学统计软件*	A. SPSS 5(8.20)	B. SAS 58(95.08)	C. Stata 9(14.75)	D. R语言 34(55.74)	E. Python 9(14.75)
2. 学习医学统计软件的难度	A. 非常困难 16(26.23)	B. 比较困难 36(59.02)	C. 一般 9(14.75)	D. 比较简单 0(0.00)	E. 非常简单 0(0.00)
3. 对R语言的了解程度	A. 完全不了解 13(21.31)	B. 基本不了解 21(34.43)	C. 一般 23(37.70)	D. 比较了解 4(6.56)	E. 非常了解 0(0.00)
4. 教学上的困惑(倾向解决途径*	A. 询问同学 52(85.25)	B. 请教教师 31(50.82)	C. 搜索引擎 56(91.80)	D. 查阅文献 14(22.95)	E. 其他 1(1.64)
5. 这些方式能否很好解决问题	A. 完全解决 2(3.28)	B. 基本解决 27(44.26)	C. 一般 29(47.54)	D. 基本没解决 2(3.28)	E. 完全没解决 1(1.64)
6. 其中最有效的是	A. 询问同学 14(22.95)	B. 请教教师 13(21.31)	C. 搜索引擎 33(54.10)	D. 查阅文献 0(0.00)	E. 其他 1(1.64)
7. 软件操作是否时常遇到问题	A. 总是 16(26.23)	B. 经常 36(59.02)	C. 偶尔 8(13.11)	D. 有时 1(1.64)	E. 几乎没有 0(0.00)
8. 软件操作问题(倾向解决途径*	A. 询问同学 43(70.49)	B. 请教教师 21(34.43)	C. 搜索引擎 50(81.97)	D. 查阅文献 6(9.84)	E. 其他 2(3.28)
9. 这些方式能否很好解决问题	A. 完全解决 3(4.92)	B. 基本解决 29(47.54)	C. 一般 27(44.26)	D. 基本没解决 1(1.64)	E. 完全没解决 1(1.64)
10. 其中最有效的是	A. 询问同学 17(27.87)	B. 请教教师 15(24.59)	C. 搜索引擎 27(44.26)	D. 查阅文献 1(1.64)	E. 其他 1(1.64)
11. 对LLM的了解程度	A. 非常了解 2(3.28)	B. 比较了解 17(27.87)	C. 一般 35(57.38)	D. 基本不了解 5(8.20)	E. 完全不了解 2(3.28)
12. 对LLM答复正确性的态度	A. 完全相信 1(1.64)	B. 基本相信 27(44.26)	C. 一般 26(42.62)	D. 比较质疑 7(11.48)	E. 非常质疑 0(0.00)
13. 对LLM帮助学习的态度	A. 完全相信 4(6.56)	B. 基本相信 43(70.49)	C. 一般 14(22.95)	D. 比较质疑 0(0.00)	E. 非常质疑 0(0.00)
14. 对LLM提升学习兴趣的态度	A. 完全相信 4(6.56)	B. 基本相信 32(52.46)	C. 一般 22(36.07)	D. 比较质疑 2(3.28)	E. 非常质疑 1(1.64)
15. LLM预期解决问题*	A. 基础理论知识的学习 45(73.77) B. 软件选择问题 35(57.38) C. 软件安装与配置问题 37(60.66) D. 命令与编程问题 49(80.33) E. 数据整理 45(73.77) F. 数据分析 46(75.41) G. 结果解释 40(65.57)				

注:*表示多选题。



2.1.1 选课学生学习基础

关于对常用医学统计软件的接触和了解程度,针对教学重点R语言,仅有34名学生接触过R(55.74%),大部分学生不了解(55.74%)。自评学习难度,52名学生认为学习医学统计软件“比较困难”或“非常困难”(85.25%)。

关于问题解决途径,对于理论知识的疑问,绝大部分学生采取的方式为“询问同学或朋友”(85.25%)与“搜索引擎”(91.80%),仅有29名学生表示传统方式解决了问题(47.54%);对于统计软件操作,85.25%的学生表示经常遇到问题,大家常用的解决办法是“询问学生”(70.49%)和“搜索引擎”(81.97%),只有52.46%的学生表示传统方式解决了问题。

2.1.2 选课学生对LLM的认识和课程期待

大多数学生对LLM的了解“一般”(57.38%),少部分学生“比较了解”(27.87%),总体了解程度有限。学生对LLM答复的信任度以“基本相信”(44.26%)

和“一般”(42.62%)为主。在LLM的作用态度上,学生对其的信任度和期待较高。77.05%的学生相信LLM助于学习医学统计软件;59.02%的学生认为LLM有助于提升学习兴趣。

对LLM预期解决问题方面学生呈积极态度,期待解决的问题每项都超过50%的学生选择,尤其是在编程(80.33%)、数据分析(75.41%)和理论学习方面(73.77%)。

2.2 实施效果

课程结束后,从课程的总体满意度、对LLM辅助教学的反馈、学习反思情况等多个角度展开问卷调查,见表2。75名学生完成课程学习,共60人填写问卷,剔除逻辑错误问卷1份,问卷有效率为98.33%。教学团队规定最终考评成绩 ≥ 70 分合格,80~89分为良好,成绩 ≥ 90 分为优秀。结果显示,班级平均成绩为81.52分,有82.66%的学生考查成绩达到良好及以上。

表2 “常用统计软件应用”教改效果调查

题目	选项[n(%)]				
	非常满意	比较满意	一般	比较不满意	非常不满意
1. 对课程总体满意度	12(20.34)	22(37.29)	20(33.90)	4(6.78)	1(1.69)
2. 对教学内容满意度	13(22.03)	23(38.98)	18(30.51)	4(6.78)	1(1.69)
3. 教学上的困惑,倾向解决途径*	A. 询问同学 42(71.19)	B. 请教教师 25(42.37)	C. 搜索引擎 44(74.58)	D. 查阅文献 13(22.03)	E. LLM28 (47.46)
4. 软件操作问题,倾向解决途径*	A. 询问同学 39(66.10)	B. 请教教师 20(33.90)	C. 搜索引擎 44(74.58)	D. 查阅文献 15(25.42)	E. LLM31 (52.54)
5. 最倾向的答疑方式▲	A. 询问同学 12(20.34)	B. 请教教师 19(32.20)	C. 搜索引擎 13(22.03)	D. 查阅文献 1(1.69)	E. LLM14 (23.73)
6. LLM实际发挥的作用*	A. 基础理论知识的学习 45(76.27) B. 软件选择问题 27(45.76) C. 软件安装与配置问题 32(54.24) D. 命令与编程问题 41(69.49) E. 数据整理 26(44.07) F. 数据分析 26(44.07) G. 结果解释 28(47.56)				
7. 学习医学统计软件的难度	A. 非常困难 24(40.68)	B. 比较困难 22(37.29)	C. 一般 11 (18.64)	D. 比较简单 2(3.39)	E. 非常简单 0(0.00)
8. 对LLM提升学习兴趣的态度	6(10.17)	24(40.68)	25(42.37)	3(5.08)	1(1.69)
9. 对LLM提升学习主动性的态度	4(6.78)	26(44.07)	24(40.68)	5(8.47)	0(0.00)
10. LLM帮助学习的态度	5(8.47)	27(45.76)	21(35.59)	6(10.17)	0(0.00)
11. LLM对增强实践意识和软件实操能力的效果	3(5.08)	29(49.15)	22(37.29)	5(8.47)	0(0.00)
12. 验证LLM回答的准确性	A. 是 35(59.32)		B. 否 24(40.68)		



续表

题目	选项[n(%)]				
	非常满意	比较满意	一般	比较不满意	非常不满意
13. 验证 LLM 准确性的方式*	A. 查找课本、PPT 等教材内容 26(44.07) B. 请教教师 20(33.90) C. 搜索引擎(如百度、谷歌等)20(33.90) D. 查阅相关文献 11(18.64) E. 多个大语言模型相互验证 15(25.42)				
14. 对 LLM 答复正确性的态度	3(5.08)	35(59.32)	20(33.90)	1(1.69)	0(0.00)
15. LLM 回复的丰富度	6(10.17)	34(57.63)	17(28.81)	1(1.69)	1(1.69)
16. LLM 的优势*	A. 即时反馈 49(83.05) B. 交互性和个性化回复 36(61.02) C. 不受时间区域限制 35(59.32) D. 知识资源丰富 37(62.71) E. 方便获得 39(66.10)				
17. LLM 的劣势*	A. 产生没有证据支持的信息 26(44.07) B. 提供的回答,缺乏详细描述 31(52.54) C. 会给出令人信服的错误信息 34(57.63) D. 只能根据预先训练过的过去知识提供答案 29(49.15) E. 互动形式局限 22(37.29) F. 其他——无 1(1.69)				
18. 使用 LLM 的最佳时间	A. 授课之前 25(42.37)	B. 课程中 22(37.29)	C. 授课之后 11(18.64)	D. 课后复习时 2(3.39)	
19. 是否愿意继续使用 LLM 辅助学习	9(15.25)	31(52.54)	17(28.81)	2(3.39)	0(0.00)
20. 是否愿意推荐该 LLM 给同学	11(18.64)	29(49.15)	19(32.20)	0(0.00)	0(0.00)

注:*表示多选题;▲表示排序题。

2.2.1 LLM增加了学生解决问题的途径

基于前期问卷调查,经过课程后,在学习中对需解答的理论问题,选择 LLM 的学生增加($P<0.001$),选择搜索引擎的人数减少($P=0.022$);遇到实践难题,选择文献学习($P=0.045$)和 LLM 途径($P<0.001$)的人数均增加,见表 3。所有途径中,请教教师是学生最倾向的解决途径(32.20%),LLM 次之(23.73%),见表 2。

2.2.2 LLM在课程学习中的实际作用

课程学习后,仍有较多学生认为学习医学统计软件困难($P=0.510$),但人数从 85.25% 降至 77.97%。学生们对 LLM 帮助学习的态度转变明显($P=0.018$),认同有帮助的人数从原来的 77.05% 下降至 54.23%,见表 3。50.85% 的学生对提升学习主动性的效果比较满意;大部分的学生认为 LLM 主要帮助理论知识学习(76.27%)和软件代码编写(69.49%),见表 2。

2.2.3 学生对 LLM 的辩证思考

LLM 激发了学生的批判性思考,59.32% 的学生会主动采用其他形式验证 LLM 的回答准确性,其中 44.07% 的学生会更信赖教材内容,见表 2。同时发现,LLM 一个很大的劣势是会给出令人信服的错误信息(57.63%)。但与课前相比,LLM 回答的准确性得到更多学生的认可($P=0.016$),见表 3。67.80% 的学生认为其回答内容较为丰富,对于使用 LLM 的优势主要集中在及时反馈(83.05%)和方便获得(66.10%)两点,见表 2。

经过调查,从表示“比较满意”和“非常满意”的结果来看,课程总体满意度(34 人,57.63%)和教学内容满意度(36 人,61.05%)反馈较好。40 人(67.79%)表示“非常愿意”和“比较愿意”继续使用 LLM 辅助学习和推荐给其他同学,上课前(42.37%)和课上(37.29%)是学生认为使用 LLM 的最佳时间,见表 2。



表 3 课程前后对待 LLM 的行为态度改变

变量	课前 $n(\%)$	课后 $n(\%)$	χ^2/W	P
理论难题寻求帮助的途径*				
询问同学	52(85.25)	42(71.19)	2.714	0.100
请教教师	31(50.82)	25(42.37)	0.554	0.457
搜索引擎	56(91.80)	44(74.58)	5.228	0.022
查阅文献	14(22.95)	13(22.03)	0.000	1.000
LLM	1(1.64)	28(47.46)	31.901	<0.001
实践困难寻求帮助的途径*				
询问同学	43(70.49)	39(66.10)	0.103	0.749
请教教师	21(34.43)	20(33.90)	0.000	1.000
搜索引擎	50(81.97)	44(74.58)	0.579	0.447
查阅文献	6(9.84)	15(25.42)	4.026	0.045
LLM	2(3.28)	31(52.54)	34.079	<0.001
学习统计学软件的难度变化			1 915.500	0.510
非常困难	16(26.23)	24(40.68)		
比较困难	36(59.02)	22(37.29)		
一般	9(14.75)	11(18.64)		
比较简单	0(0.00)	2(3.39)		
非常简单	0(0.00)	0(0.00)		
LLM 帮助学习的态度			2 198.500	0.018
完全无效	0(0.00)	0(0.00)		
基本无效	0(0.00)	6(10.17)		
一般	14(22.95)	21(35.59)		
比较有效	43(70.49)	27(45.76)		
非常有效	4(6.56)	5(8.47)		
LLM 提升学习兴趣的态度			1 864.000	0.714
完全无效	1(1.64)	1(1.69)		
基本无效	2(3.28)	3(5.08)		
一般	22(36.07)	25(42.37)		
比较有效	32(52.46)	24(40.68)		
非常有效	4(6.56)	6(10.17)		
LLM 回答准确性的态度			1 386.500	0.016
非常不同意	0(0.00)	0(0.00)		
不同意	7(11.48)	1(1.69)		
一般	26(42.62)	20(33.90)		
同意	27(44.26)	35(59.32)		
非常同意	1(1.64)	3(5.08)		

注: *表示多选题转换为二分类变量采用卡方检验分析,统计量为 χ^2 ;其余题目选项为有序分类变量,采用秩和检验分析,统计量为 W 。

3 讨论

课程考核成绩和调查结果显示,本次教学改革的实施成效良好,学生整体满意度较高,倾向使用 LLM 解决问题。且学生在信任 LLM 技术准确性的同时,对学习帮助价值评价更加谨慎,这种审慎态度的形成是学生开始建立风险意识的积极表现。最重要的是,59.32% 的学生养成了主动验证的习惯,44.07% 回归教材求证,体现了学生批判性思维的重要发展。在

课程结束后,虽然主观上认为统计软件学习困难的学生数量下降并不显著,但 50% 以上学生认可 LLM 在提升学习主动性和实践能力方面的价值。本研究中 LLM 在学习行为和学习态度方面的积极影响,与陈东河等人的研究结论相一致^[17]。此外,研究表明,改革后的 LLM 教学融合培养了学生理性使用 LLM 的素养,有效促进了学生的思辨能力,降低学生自主探索的风险,预防 LLM 过度依赖。



LLM融入课程学习虽有其必然优势,但也存在局限性。LLM为学生提供了几个潜在优势,包括直接访问大量信息、没有时域限制、对话式交流促进个性化学习体验以及增强常用医学统计学软件应用实践技能发展。对于教师来说,LLM可以促进教师教授复杂统计学概念和软件应用,唤起学生实践积极性,并减少教育工作者的工作量。但其互动形式局限、理解准确性和回答可靠性有待提高,过度依赖可能给学生的学习带来误导,这与很多研究的担忧相同。在保健教育学和医学研究生使用中,作者提到LLM幻觉概念、产生虚拟信息、缺乏科学准确性、更新知识有限以及缺乏批判性地讨论结果的能力^[18-19]。关于运用较广的ChatGPT临床中教育中的应用研究发现^[20],LLM虽可协助构建逼真的临床模拟场景,提高教学质量,提高学生的实践技能,但存在知识过时和不准确之处的问题。故将LLM融入课程教学,更需要教育者进行系统化设计和正向引导,教师的核心职责应从单纯的知识传授,转向培养学生成为LLM输出的“合格质检员”。

本研究也存在几点不足,由于问卷收集中匿名设计使研究无法进行个体层面的配对分析,可能低估了教改的干预效果;此外,本研究旨在评价LLM主动引入课题后的学习态度变化和对LLM的认识改变,在教学效果量化方面,由于既往课程未收集可比的问卷数据,同时不同学期的教学内容与学生构成存在差异,因而无法开展严格的历史对照。未来的教学改革效果研究可开展平行对照,更好地比较传统教学与LLM辅助教学的效果。另外,由于学生较多使用LLM解决实践中的编码问题,现实中编码问题种类繁多,学生对LLM输入的表达也存在偏差,而AI的准确性取决于其训练数据的质量和多样性^[21],后续可进一步对“医学统计助教”模型进行改良,提升算法,并优化更新资料库以满足多样化需求。

在数智化时代,LLM不可避免地成为被大众广泛使用的工具,批判性评估AI生成内容的能力已然成为当代大学生应具备的一项基本能力。未来更应积极将LLM使用教育融入教学,正确引导学生如何使用LLM,包括告知其局限性和优势,指导学生辨别AI提供信息的真实性和准确性,以及遵守伦理道德标准^[22-23]。这样,我们才能将LLM的局限性转化为宝贵的学习资源,最终培养出既能利用技术赋能、又能驾驭技术风险的新一代学者。教育界也迫切需要制定LLM在科学研究和课程学习中的使用指南,同时考虑问责制、诚信、透明度和诚实度等问题^[24]。

此外,在集成专业领域的微调GAI也体现出其价值和前景,专业化的AI将提供更加准确和贴合教学的输出,方便随时随地解答教学中学生的个性化问题,破除了传统寻求教师帮助的时域局限性。同时,固定LLM的使用,教师可以在课前验证其安全性、可靠性和可重复性,并评估其对教学内容和质量的影响,以防止对教学过程产生不利影响^[20,23,25]。

参考文献:

- [1]曹慧芬.大数据背景下R语言在医学统计学教学实验中的应用[J].科教导刊,2021(25):149-151.
- [2]陈玉云,欧阳东,王耀民,等.基于R语言的统计分析在医学类专业的教学实践与探索[J].科技资讯,2024,22(20):220-223.
- [3]阎洁,杨俊丽,王建文,等.R语言在医学院校生物信息学实验教学中的应用与探索[J].医学信息学杂志,2020,41(1):87-89.
- [4]LIU J L, LIU F, FANG J B, et al. The application of Chat Generative Pre-trained Transformer in nursing education[J]. Nursing Outlook, 2023, 71(6):102064.
- [5]TANGADULRAT P, SONO S, TANGTRAKULW-ANICH B. Using ChatGPT for clinical practice and medical education: Cross-sectional survey of medical students' and physicians' perceptions[J]. JMIR Medical Education, 2023(9):e50658.
- [6]李洪修,李晓漪.数智化背景下学校课堂教学形态的转型与实现[J].教育理论与实践,2024,44(34):11-18.
- [7]孙文文,张歆翊,王若桐,等.AI赋能空气理化检验课程教学应用[J].医学教育研究与实践,2025,33(5):682-687.
- [8]KUNG T H, CHEATHAM M, MEDENILLA A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models[J]. PLoS Digital Health, 2023, 2(2):e0000198.
- [9]DAI H X, LIU Z L, LIAO W X, et al. AugGPT: Leveraging ChatGPT for text data augmentation[J]. IEEE Transactions on Big Data, 2025, 11(3):907-918.
- [10]OPENAI, ACHIAM J, ADLER S, et al. GPT-4 technical report[EB/OL]. (2023-03-15) [2025-10-22]. <https://arxiv.org/abs/2303.08774>
- [11]PARK J. Medical students' patterns of using ChatGPT as a feedback tool and perceptions of ChatGPT in a Leadership and Communication course in Korea: A cross-sectional study[J]. Journal of Educational Evaluation for Health Professions, 2023(20):29.
- [12]SIMMS R C. Work with ChatGPT, not against: 3 teach-



- ing strategies that harness the power of artificial intelligence[J]. Nurse Educator, 2024, 49(3):158-161.
- [13] SHUE E, LIU L, LI B X, et al. Empowering beginners in bioinformatics with ChatGPT[J]. Quantitative Biology, 2023, 11(2):105-108.
- [14] LI Y J, CHOI D, CHUNG J, et al. Competition-level code generation with AlphaCode[J]. Science, 2022, 378(6624):1092-1097.
- [15] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. December 6-12, 2020, Vancouver, BC, Canada. ACM, 2020:1877-1901.
- [16] CHOWDHERY A, NARANG S, DEVLIN J, et al. PaLM: Scaling language modeling with pathways[J]. Journal of Machine Learning Research, 2023, 24(1):11324-11436.
- [17] 陈冬河, 杨君, 翁婉雯, 等. 人工智能辅助非影像专业住院医师核医学PET短期培训探索[J]. 医学教育研究与实践, 2025, 33(5):738-743.
- [18] SALLAM M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns[J]. Healthcare, 2023, 11(6):887.
- [19] 李海琛, 王臻, 马婷, 等. 生成式人工智能辅助医学研究生开题设计路径与反思[J]. 医学教育研究与实践, 2025, 33(4):492-500.
- [20] XU X J, CHEN Y X, MIAO J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: A systematic scoping review[J]. Journal of Educational Evaluation for Health Professions, 2024(21):6.
- [21] ALMAZYAD M, ALJOFAN F, ABOUAMMOH N A, et al. Enhancing expert panel discussions in pediatric palliative care: Innovative scenario development and summarization with ChatGPT-4[J]. Cureus, 2023; 15(4): e38249.
- [22] LENIHAN D. Three effective, efficient, and easily implementable ways to integrate A. I. into medical education[J]. Cureus, 2023, 15(10):e47204.
- [23] ABD-ALRAZAQ A, ALSAAD R, ALHUWAIL D, et al. Large language models in medical education: Opportunities, challenges, and future directions[J]. JMIR Medical Education, 2023(9):e48291.
- [24] VAN DIS E A M, BOLLEN J, ZUIDEMA W, et al. ChatGPT: Five priorities for research[J]. Nature, 2023, 614(7947):224-226.
- [25] KARABACAK M, OZKARA B B, MARGETIS K, et al. The advent of generative language models in medical education[J]. JMIR Medical Education, 2023(9):e48163.
- (编辑: 赵伊昕)