多智能体协同的医学Python数据分析实验辅导系统设计与实践

梁 洁[1],石 林[1],温川雪[2],毕志升[2]
广州医科大学:1.信息与数据管理中心;2.生物医学工程学院,广州 511436

【摘要】新医科建设背景下,为解决医学院校Python数据分析实验教学中编程记忆检索难、调试纠错慢、概念迁移难、任务拆解乱等问题,基于教学实情和自我效能感等理论指导,设计一个基于本地化部署的多智能体协同实验辅导系统,实现医学术语解释、医学类比讲解、代码诊断及任务分解等核心功能,并构建贯穿课前、课中、课后全流程的实验教学支撑体系。教学实践表明,系统意图识别准确率92.7%、回答匹配度95.8%、整体满意度4.0分(5分制)。试验组课堂实验完成率(89.2%)高于对照组(71.9%),课后综合实践完成率(93.2%)高于对照组(84.4%),试验组编程自我效能感提升(1.07分)亦高于对照组(0.36分)。试验组的教师答疑人次较对照组下降60%,基础性提问减少18.6%,试验组个性化指导时间提升至48%。系统在识别准确率、回答匹配度及用户满意度方面均表现良好,显著提升了师生双向效能,为医学编程类实验课程的智能化辅导提供了有益参考。

【关键词】多智能体;大语言模型;医学Python数据分析;实验教学辅导
【中图分类号】G642;R-4 【文献标志码】A 【文章编号】1004-5287(2026)05-0653-07
【DOI】10.13566/j.cnki.cmet.cn61-1317/g4.202605010

Design and practice of a multi-agent collaborative experimental tutoring system for medical Python data analysis

LIANG Jie[1], *SHI Lin*[1], *WEN Chuanxue*[2], *BI Zhisheng*[2]
Guangzhou Medical University: 1. *Information and Data Management Center*;
2. *School of Biomedical Engineering*, *Guangzhou* 511436, *China*

【**Abstract**】Under the background of New Medicine development, this study aims to solve the problems in Python data analysis experimental teaching in medical colleges, including difficulties in programming memory retrieval, inefficient debugging and error correction, obstruction in concept transfer, and disorganized task decomposition. Based on teaching practices and theoretical guidance, including self-efficacy theory, this study designs and implements a locally deployed multi-agent collaborative experimental tutoring system that realizes core functions such as medical terminology explanation, medical analogy-based instruction, code diagnosis, and task decomposition, and constructs a comprehensive experimental teaching support system covering the entire process of pre-class, in-class, and after-class activities. Teaching practice demonstrates that the system achieves an intent recognition accuracy of 92.7%, a response relevance rate of 95.8%, and an overall satisfaction score of 4.0 out of 5.0. The experimental group's in-class experiment completion rate (89.2%) was higher than that of the control group (71.9%), while the after-class comprehensive practice completion rate

基金项目:2025年度广东省本科高校教学质量与教学改革工程建设项目专项人才培养计划(256);广州市教育科学规划2026年度课题项目(2026115095);广州医科大学2024年度本科教学质量与教学改革工程建设项目(2024JXGG11)
收稿日期:2026-05-06
作者简介:梁洁,高级工程师,硕士,研究方向为教育技术。
通信作者:毕志升,副教授,博士,研究方向为计算机应用技术。E-mail:bivictor@gmail.com

(93.2%) exceeded that of the control group (84.4%). The improvement of programming self-efficacy in the experimental group (1.07) was also higher than that of the control group (0.36). In the experimental group, the number of teacher Q&A sessions decreased by 60% compared with the control group, basic questions decreased by 18.6%, and the proportion of individualized guidance time increased to 48%. The system performed well in recognition accuracy, answer matching, and user satisfaction, significantly improving the bidirectional efficiency between teachers and students and providing a useful reference for intelligent tutoring in medical programming laboratory courses.

【Keywords】 multi-agent; large language model; medical Python data analysis; experimental teaching tutoring

随着医学进入大数据与精准医疗时代,医工交叉融合对医学生的数据素养和编程能力提出了更高要求。国务院办公厅《关于加快医学教育创新发展的指导意见》(国办发〔2020〕34号)明确提出,以新医科统领医学教育创新,推进医科与理科、工科、文科等多学科深度交叉融合,加快培养“医学+X”高层次复合型医学人才[1]。在此背景下,Python数据分析作为医学院校面向大一学生开设的计算机基础通识课,在课程体系中的地位日益凸显。课程旨在培养医学院校学生的编程实践能力与Python医学数据分析能力,为后续专业学习与科研实践奠定计算机基础。然而,该课程实验教学中普遍存在学生编程记忆检索难、调试纠错慢、概念迁移转换难、任务拆解乱等问题。受大班教学与有限课时的双重制约,教师的个性化指导在覆盖面与深入度上难以兼得,往往只能顾及部分学生,无法有效回应全体学生的差异化学习需求。

近年来,生成式人工智能与智能体技术的快速发展为实验教学改革提供了新路径。已有研究表明,大语言模型(large language model, LLM)驱动的智能辅导系统在代码调试[2-3]、个性化答疑[4-5]、项目式学习[6-7]等方面展现出显著成效。因此,本研究设计了一个基于本地化部署的多智能体协同实验辅导系统,实现了医学解释、知识讲解、代码调试、任务分解四类智能体协同工作,旨在解决医学院校Python数据分析实验教学中的问题,为医学交叉学科背景下的实验教学智能化辅导提供系统设计参考与实现方案。

1 Python数据分析实验教学的核心问题

Python数据分析课程聚焦于Python语言在数据处理与分析领域的核心应用,而Python语言已成为当前数据科学领域主流的编程实践。学习本课程前,学生已修读一学期的Python基础知识,掌握了基本语法与编程逻辑,为本课程的教学开展奠定了基础。然而,从基础学习迈向医学数据的应用实践,仍存在显著的能力鸿沟。经过近几年的实验教学分析,学生实验存在4方面学习问题。①函数方法记忆与检索难,学习效能感低。主要体现在库函数数量多且命名规则不统一、记忆负担重,学生仅学过基础语法,对第三方库调用模式不熟悉,函数讲解与医学应用脱节,难以建立关联记忆。②代码错误诊断难,修复效率低。主要在于数据类型不匹配、缺失值传播等错误的隐蔽性强,学生难以定位问题根源。③医学概念与编程方法的转换难,课程价值认同感弱。学生理解医学指标含义却不知如何用代码表达,能输出分析结果却无法从医学角度解读。④实验项目规划与拆解乱,任务完成效能不足。在面对综合性实验项目时,学生不知从何入手,难以合理拆解任务。

2 人工智能与智能体技术的教育应用

以ChatGPT、DeepSeek等为代表的大语言模型的突破性进展,推动了人工智能从工具向智能体的演进[8]。智能体系统具备自主感知、任务规划、工具调用与协同执行等能力,能够模拟人类教师的辅导行为,在教育场景中发挥独特优势[9]。在编程教育领域,智能体可承担代码纠错、概念讲解、学习路径推荐等任务,有效缓解教师重复性答疑压力[10]。在医学教育领域,基于智能体的虚拟患者[11]、临床推理训练系统[12]等已初显成效。然而,将智能体技术系统性地应用于医学编程实验教学、医学数据分析课程,仍缺乏可参考的实践方案。

尽管大语言模型与智能体技术在教育领域展现出巨大潜力,但仍存在一些局限。从大语言模型角度看,通用大模型依赖于静态知识训练,在特定学科专业领域易产生“幻觉”回答[13],且回答内容常超出课程边界,难以为教学辅导所直接使用。从单智能体角度看,单一智能体难以同时承担多种任务角色,且边界模糊,响应质量不稳定。据此,检索增强生成(retrieval-augmented generation, RAG)技术通过动态调用外部知识库,可有效提升回答的准确性与领域适配性[14]。随着智能体教育应用的推广,智能体逐渐从单一向多元化发展[15]。多智能体系统具有任务并行处理能力,通过角色分工与协同机制能够完成复杂度高的教学任务,提升专业性与响应精度[16-17]。

3　多智能体协同的医学 Python 数据分析实验辅导系统

为解决上述 Python 实验教学的核心问题，本研究依据自我效能感理论进行系统设计。该理论指出，自我效能感取决于个体对任务所需技能与资源的评估，而直接的成功经验是提升效能感的最有效途径[18]。然而，要有效落实这一理论机制，单一技术手段难以全面应对。鉴于此，本研究融合检索增强生成、本地知识库与多智能体协同技术，设计了一个面向医学 Python 数据分析实验教学的多智能体协同辅导系统。

3.1　系统架构

系统基于本地化部署架构，融合 LLM、RAG、意图识别与多智能体协同等技术，构建了统一入口、智能分发、专业处理、统一反馈的辅导系统，其架构如图 1 所示。系统采用 B/S 架构，前端提供对话交互界面，后端以本地部署的 LLM（Qwen3-235B-A22B-Thinking-2507-W8A8）为认知推理核心。学生通过统一入口提交提问后，系统首先通过预设的规则分类与关键词匹配机制进行意图识别，判定问题类型（医学概念解释、函数方法查询、代码错误诊断、项目任务拆解），再将请求分发给对应的功能智能体。各智能体通过 RAG 技术动态调用本地知识库（涵盖医学案例、课程资料、常见错误、实验题集等）。知识库基于 Qdrant 向量数据库存储，文档分块后由 bge-m3 模型完成向量化。系统检索相关文本片段后，将其与原始问题拼接，提交 LLM 生成回答，最终由入口模块统一汇聚各智能体输出后反馈给学生。

系统部署于学校本地算力服务器，采用 2 台计算节点＋1 台管理节点的部署架构。单台计算节点搭载 2 颗鲲鹏 920 处理器（主频 2600 MHz，48 核）、8 张昇腾 910B NPU（总显存 512 GB）及 1024 GB 内存，主要承担 LLM 推理以及 RAG 检索任务；单台管理节点配置 2 颗鲲鹏 920 处理器（主频 2600 MHz，48 核）、2 张 Atlas 300I Duo 加速卡（总显存 192 GB）及 512 GB 内存，负责集群任务调度与系统监控。LLM 模型采用 W8A8 量化方案部署，基于昇腾 910B NPU 集群算力完成推理运行，在复杂代码推理与医学语义理解任务中兼顾了性能与资源效率。所有推理与知识检索均在校园网内部完成，保障数据安全与响应时效。

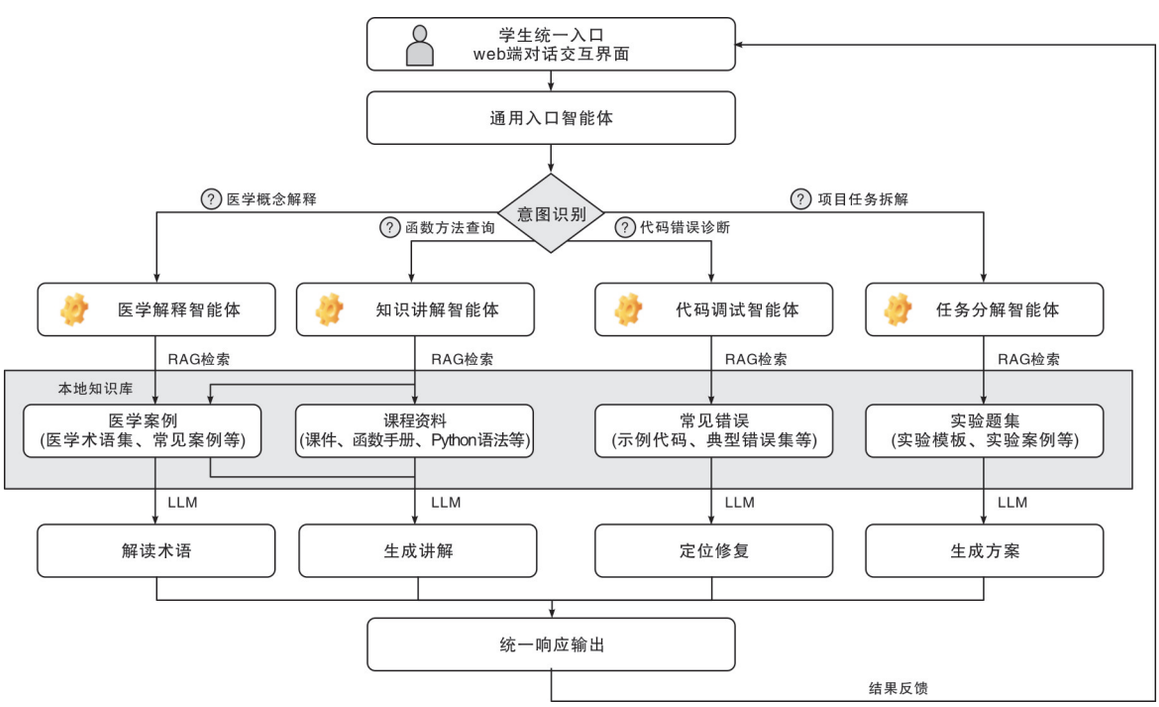


图 1　医学 Python 数据分析实验辅导系统架构设计图

3.2　智能体功能设计

针对上述 Python 实验教学的核心问题，学生提问可归纳为医学概念解释、函数方法查询、代码错误诊断、项目任务拆解 4 类。在此基础上，设计相应问题解决的智能体，加上通用入口智能体，本系统共设计 5 个智能体。①通用入口智能体：作为系统交互中枢，负责接收学生自然语言输入，解析问题类型、分发至对应智能体，并整合处理结果统一反馈。学生只需自然提问，无须自行分类。②医学解释智能体：专注于统计指标与医学术语的临床意义解读，旨在帮助学生建立数据分析结果与医学问题之间的关联。接收输入后，经 RAG 检索医学案例库，由 LLM 生成面向

临床的解读内容。作为本系统的特色模块，该智能体区别于通用编程辅导系统，其核心价值在于教会学生用代码表达医学意义，将数字转化为临床价值。③知识讲解智能体：负责处理函数方法查询问题。其核心任务是讲解函数用法与参数，结合医学案例演示，提供代码示例与练习建议。收到提问后经RAG检索文档，按“功能-参数-案例-代码”格式生成回答，建立函数与医学场景的关联记忆。④代码调试智能体：负责处理代码错误诊断问题。其核心任务是分析代码与错误信息，定位错误类型及位置，生成符合教学进度的修复建议与修改代码。收到报错后，以解析错误、RAG检索、LLM生成的顺序执行。遵循不超纲、不代写原则，仅修复错误部分并附带教学注释，避免引入未学语法。⑤任务分解智能体：负责处理项目任务拆解问题。其核心任务是分析实验项目需求，识别数据探索、清洗、统计分析、可视化呈现等项目阶段，拆解项目任务步骤，并给予核心函数与实践建议。智能体收到需求后，先从RAG检索模板库匹配相似案例，再由LLM结合需求生成个性化方案。

以上各智能体角色定位清晰、功能边界明确，通过协同工作覆盖学生实验辅导的全流程需求，针对不同实验需求提供差异化的精准智能支持。其中，医学解释智能体补齐知识评估短板，知识讲解智能体通过医学类比降低记忆负担，代码调试智能体提供即时修复反馈以积累成功体验，任务分解智能体帮助学生清晰拆解任务结构。通过降低认知负荷、提供即时反馈、增强成功体验，从根源上缓解学生实验压力、消除学习障碍，提升学生的编程自我效能感。

3.3 知识库构建

本地知识库是本系统回答准确性的核心保障。首先，保障回答准确性，基于课程专属知识生成回答，避免通用LLM的幻觉问题；其次，实现教学同步，将隐性教学经验（如常见错误、典型案例）显性化为可复用的知识资产；最后，突出医学特色，医学案例库确保系统回答具有医学专业性，区别于通用编程辅导工具。

知识库由医学案例库、课程资料库、常见错误库、实验题集库组成。①医学案例库：该库是区别于通用编程辅导系统的特色模块，由医学术语集、医学指标集和常见医学案例三部分组成，旨在补足LLM在医学场景下的理解能力。所有医学指标均参考最新临床指南，如《中国高血压防治指南（2024年修订版）》等，案例内容均经临床医学专业教师审核把关。各组成部分的核心内容如表1所示。该库主要服务于医学解释智能体和知识讲解智能体，用于统计指标的临床解读和函数方法的医学场景类比。②课程资料库：该库是知识库的基础模块，涵盖课程教学过程中产生的各类数字化资源。具体包括课程课件（各章节教学PPT与讲义）、函数方法手册（pandas、matplotlib等库的常用函数说明与参数详解）、Python语法参考（基础语法规则与代码规范）等。③常见错误库：该库重点提取历届学生在实验过程中出现的典型错误代码、错误类型及修复方案，增强语料的问题导向性。④实验题集库：该库是综合性实验项目规划与拆解的任务模板库。采用基于实验模板的实验任务书与项目案例，将门诊流量分析、体检指标统计、心脑血管风险分析等医学案例转化为结构化实验任务描述。

表1 医学案例库的核心内容

类别	内容说明	典型示例
医学术语集	医学专业术语的定义与解释	收缩压/舒张压、体质指数（BMI）、*P*值、置信区间、灵敏度、特异度
医学指标集	常见指标的参考范围与异常筛查阈值	血压（正常＜120/80 mmHg，高血压筛查≥140/90 mmHg）、血糖（正常空腹3.9～6.1 mmol/L，糖尿病筛查≥7.0 mmol/L）、血脂（正常＜5.2 mmol/L，高血脂筛查≥6.2 mmol/L）
常见医学案例	典型医学数据分析场景，配套数据、代码及解读	门诊流量分析、体检数据处理、疾病风险预测、药品数据分析

为提升知识库检索精度及优化信息呈现方式，系统依据认知负荷理论进行知识库建设。该理论指出，减少冗余信息呈现、降低外在认知负荷，有助于提升学习效果[19]。为此，在检索精度提升上，对语料进行语义、递归及固定字符大小相结合的智能分段，即先按语义边界初步切分，再对超长文本递归拆分，最后以固定字符大小为上限，切分成小粒度知识片段，控制知识单片段信息量在记忆容量之内，让检索结果更加精确。在检索结果信息呈现上，系统在回复中同步显示所引知识片段来源，学生可点击查看上下文，仅呈现与当前问题直接相关的内容，避免无关信息占用认知资源，从而降低学生的外在认知负荷。同时，依据最近发展区理论，系统提供知识检索与范例生成等支架支持，降低学生低效重复性任务负担[20]。通过RAG技术动态检索与当前问题最相关的知识片段，为学生提供适时的“脚手架”支持，确保辅导内容始终

处于学生可理解的区间内。

4　实验辅导系统的教学实践与效能分析

4.1　教学实践实施

本研究将多智能体协同的实验辅导系统嵌入医学院校 Python 数据分析实验教学全过程，构建了“课前知识自测—课中实验辅导—课后项目拓展”全流程协同的实验教学支撑体系，如图 2 所示。系统贯穿在课前、课中及课后各阶段，分别承担差异化辅导角色，满足学生在实验各阶段的个性化学习需求。

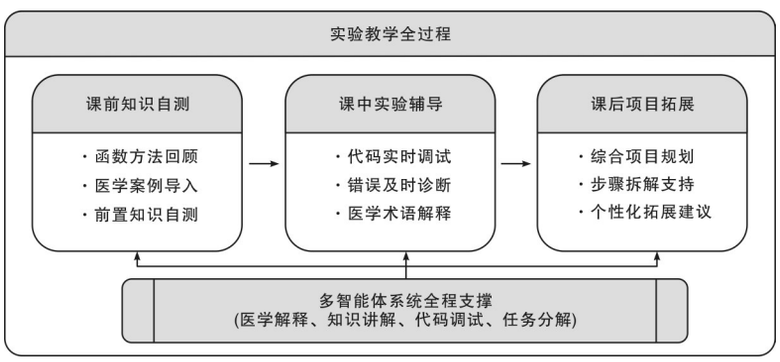


图 2　实验教学支撑体系

首先，课前知识自测环节，强调主动式、对话式的知识回顾与巩固。以医学案例为情境，将常用函数拆解为可交互的问答单元，学生在对话中提前熟悉实验课将要使用的函数，能够减少实验课开始时的茫然感。其次，课中实验辅导环节，在课堂实验过程中，学生编写代码时遇到错误，可随时通过统一入口提交代码片段。代码调试智能体实时定位错误类型与位置，生成符合教学进度的修复建议。与传统的教师巡回指导、学生排队答疑的模式不同，本阶段实现即时性、个性化的调试支持。同时，医学解释智能体针对医学术语与统计指标的疑问提供临床解读，弥补学生在数据分析结果临床转化应用方面的能力短板。最后，课后综合性实验项目拓展环节，任务分解智能体提供项目规划与步骤拆解支持。系统不会直接给学生答案，但会提供结构化、引导式的项目规划方案，将复杂项目拆解为可执行的阶段性任务，并生成代码框架，降低项目启动门槛。

该系统在 2025 — 2026 学年第二学期 Python 数据分析课程中投入使用，图 3 展示了部分多智能体系统辅导学生实验的应用示例。



图 3　实验辅导应用示例

4.2　准实验研究设计

为验证系统辅导效果，本研究采用准实验研究，设置对照组（64 名，传统教师辅导）与试验组（74 名，教师辅导＋本系统辅助）。两组均为 2025 级临床医

学专业一年级学生，由同一教师授课，教学计划与考核方式一致。Python基础前测平均成绩分别为72.6、72.9分，组间无显著差异（$P>0.05$），满足同质性要求。实验周期为8周，共28学时：理论6学时（课堂讲解NumPy、Pandas、Matplotlib）；实验16学时（课堂完成体检数据清洗、门诊流量分析等8个医学案例任务）；综合实践6学时（课后自主完成3个基于真实数据集的综合实验并线上提交）。两组唯一差异为试验组使用本系统获得即时智能辅助，对照组采用传统教师答疑。课后难以完全控制两组的工具使用行为，这是准实验设计的固有局限。因此，效能分析以课堂实验数据为基准，以综合实践完成率作为辅助参考，通过系统后台日志记录试验组实际使用情况，结合问卷数据，检验系统使用行为与综合实践完成质量之间的关联。

4.3 教学实践效能分析

4.3.1 学生实验效能分析

学生的实验效能量化分析包括：①课堂实验完成率，按学生课堂完成并提交全部8个实验任务来统计。经独立样本t检验，试验组完成率（89.2%，66/74）高于对照组（71.9%，46/64），差异有统计学意义（$P<0.01$），表明系统实时辅导有效降低了学生因调试障碍而放弃实验的概率。②综合实践完成率，按学生课后自主完成3个综合实验的报告与代码来统计。试验组完成率为93.2%（69/74），高于对照组的84.4%（54/64）（$P<0.01$）。两组线上完成率均高于线下，可能与线上实验不受时间限制、学生可自主安排有关。后台日志显示，完成全部3个综合实验的试验组学生中，67%会使用任务分解智能体规划项目。问卷分组数据显示，使用系统拆解任务的学生完成率为96.2%，高于未使用者的78.3%（$P<0.01$）。结合客观操作数据与问卷数据交叉验证可见，系统使用行为与实践完成质量存在正向关联，说明系统可有效辅助课后自主学习。③编程自我效能感分析，采用5分制Likert量表（共5题）进行前后测，涵盖任务完成信心、调试能力自信及知识迁移能力等维度，面向Python实验教学场景编制，量表Cronbach's α系数为0.86，内部信度良好。前测两组编程自我效能感均值分别为3.12和3.09分，独立样本t检验结果显示无显著差异（$P>0.05$）；后测试验组提升至4.19分（+1.07分），对照组提升至3.45分（+0.36分），独立样本t检验结果显示组间差异显著（$P<0.01$），表明系统的即时反馈与个性化辅导有效增强了学生的编程自信心。

4.3.2 教师实验教学效能分析

教师的实验教学效能量化分析包括：①教师单位课时平均答疑人次，基于课堂观察与教学日志统计。试验组为18人次，对照组为45人次，降幅达60%，表明系统有效分担了重复性答疑工作，显著缓解了大班教学中教师的课堂应答压力。②基础性提问占比，以课堂观察与教师日志（对照组）及系统意图识别模块自动分类标注（试验组）为数据来源。对照组教师接收并应答的基础性问题占其全部答疑人次的比例为52%，试验组后台日志数据统计为18.6%，此类提问由系统即时处理，不再流向教师，使教师得以将精力聚焦于高阶指导。③个性化指导时间占比，基于课堂观察与教学日志统计。试验组提升至48%，对照组不足20%。由此可见，系统将教师从重复性答疑中解放出来，使其有更多时间开展高阶思维引导，师生互动的话语层次明显提升，教师的专业判断力与教学引导力得到了更充分的发挥。

4.3.3 多智能体效能分析

课程运行期间（前8周），系统面向试验组74名学生开放，后台日志共记录对话轮次703次，提问总数2 124条。基于这些运行数据，进行多智能体效能分析，主要包括：①意图识别准确率，分析通用入口智能体对学生问题类型判断的准确性。由1名熟悉系统意图识别规则的开发人员，对2 124条提问进行逐一人工校验，以实际语义判定为基准，对系统自动分类结果进行核对与修正。系统自动分类准确率为92.7%（1968/2124），验证了意图识别的准确性。②回答与课程的匹配度，评估系统回答是否在课程知识范围内。由3名任课教师对随机抽取的500条问答记录进行独立评分（5分制，4分及以上为“匹配”）。3名教师一致评定为匹配的比例为95.8%（479/500），验证了本地知识库与RAG技术的有效性与准确性。③学生满意度，采用5分制Likert量表调查，系统整体满意度为4.0分，各智能体得分为任务分解4.2分、医学解释4.1分、知识讲解3.8分、代码调试3.5分。前两者满意度较高，反映学生对医学案例融入编程及任务规划辅助的认可；代码调试得分较低，分析显示其对链式调用逻辑错误和环境配置问题的排查能力有较为明显的不足，还需在局部定位与全局推理、诊断粒度、环境感知方面进行系统优化。此外，试验组问卷调查显示，89%的学生认为系统加深了对医学数据分析的理解，76%表示愿意继续使用，表明系统获得了学生广泛认可。

4.4 教学实践结果讨论

上述数据表明：①课堂实验完成率显著提升，课后综合实践完成率表现良好，学生编程自我效能感明

显增强，这3个维度的数据共同验证了系统通过降低认知负荷、提供支架支持和积累成功体验，在课前、课中、课后各阶段均展现出稳定的辅导效能，有效促进了学生的学习。②系统的应用优化了教师实验教学工作结构，基础性答疑占比下降，课堂应答压力缓解，教师角色从低阶应答转向高阶引导，工作体验与获得感显著提升。③意图识别准确率、回答匹配度与整体满意度3个指标共同表明，系统在准确理解学生需求、将回答限定于课程范围以及提升使用体验3个层面均表现良好，已在当前教学阶段展现出稳定的辅导效能。后续还需进一步优化代码调试智能体的复杂错误诊断能力，持续提升系统的整体辅导效能。

5　结束语

本研究以医学院校Python数据分析实验教学实情为出发点，以自我效能感理论指导系统设计，以认知负荷理论与最近发展区理论指导知识库建设，设计了一个基于本地化部署的多智能体协同实验辅导系统，实现了医学术语解释、医学类比讲解、代码错误诊断、实验任务拆解等核心功能，并将系统投入到实验教学实践中，构建了“课前知识自测—课中实验辅导—课后项目拓展”全流程协同的实验教学支撑体系，实现了与课程教学同步的实验辅导。教学实践表明，系统在优化教师实验教学工作结构的同时，有效提升了学生实验完成率与编程自我效能感，师生双向效能均得到显著改善，为医学编程类实验课程的智能化辅导提供了有益参考。

然而，本研究仍存在一些局限性：第一，系统功能不够完善，主要体现在代码调试智能体对复杂逻辑错误的排查能力有限、知识库依赖人工更新、本地化部署限制了大规模并发使用等方面；第二，准实验设计不够全面，观测周期较短，缺乏长期效果追踪；第三，系统的通用性不足，主要针对单门课程定制开发，尚未形成面向不同医学编程课程的可迁移框架；第四，评价维度偏重师生可测指标，对医工融合认同等深层成效的测评尚不充分。后续研究将针对上述四方面局限逐一优化，推动系统的功能完善、实验设计的科学性提升、课程适配性的拓展及评价体系的全面化。

参考文献

[1] 国务院办公厅. 国务院办公厅关于加快医学教育创新发展的指导意见(国办发〔2020〕34号)[EB/OL]. (2020-09-17)[2026-04-06]. https://www.nhc.gov.cn/bgt/gwywj2/202009/54f4a57e6eaf40afa48b608e4fed35b6.shtml.

[2] 王路桥，周洋涛，李青山，等. 基于大语言模型的多智能体协作代码评审人推荐[J]. 软件学报，2025，36(6)：2558-2575.

[3] 姚嘉轩，夏小俊. 大模型赋能的编程作业评估智能体设计与应用[J]. 计算机教育，2026(2)：98-103.

[4] 张力军，刘偲，廖纪童，等. 基于大模型检索增强生成的计算机网络实验课程问答系统设计与实现[J]. 实验技术与管理，2024，41(12)：186-192.

[5] 丁宁，宋雨欣，单泽田，等. 基于检索增强生成(RAG)技术的医学教学辅助智能问答系统的构建探索[J]. 中国医学教育技术，2025，39(1)：1-5.

[6] 刘雪峰，陈玮，韩旭，等. 大模型时代计算机实验教学的挑战与机遇[J]. 实验技术与管理，2025，42(8)：36-44.

[7] 王兆静，胡新荣，李丽，等. 大语言模型辅助下数据挖掘类课程教学内容与模式改革研究[J]. 计算机教育，2025(11)：74-79.

[8] 解敏，杨洁，陈佳琳，等. AI浪潮中，如何凸显“人”的价值[N]. 新民晚报，2026-03-09(3).

[9] 刘思宇，谢春晖，陶鑫，等. 基于人工智能的教育智能体建设及应用：以首都医科大学“医学英语术语”智能体建设为例[J]. 中国医学教育技术，2026，40(1)：22-27.

[10] 石蕴玉，纪雨，刘翔，等. 多智能体在面向对象编程课程设计中的应用[J]. 实验室研究与探索，2026，45(1)：153-160.

[11] 王子杰，季逸圣，郭峰. 基于人工智能模型的虚拟患者在外科住院医师培训中的应用[J]. 江苏卫生事业管理，2025，36(10)：1500-1503.

[12] 周志浩，宋佳琳，刘嘉成，等. 基于人工智能LLM技术的虚拟患者系统构建与临床教学应用[J]. 医学新知，2024，34(7)：833-842.

[13] 刘华，祝智庭. 大模型支持的会话辅导系统开发：范式转换、架构设计与能力训练[J]. 现代远程教育研究，2024，36(3)：11-19.

[14] 罗小娟，程奕豪. LightRAG框架下AI答疑系统的设计与实现[J]. 实验室研究与探索，2026，45(1)：8-14.

[15] 顾小清，郝祥军. 悟空的毫毛：正在重塑学习技术系统的多智能体[J]. 华东师范大学学报(教育科学版)，2025，43(5)：16-29.

[16] 王骁汉，江哲涵，廖珺，等. 基于多智能体的PBL模拟教学系统的构建[J]. 护理研究，2025，39(13)：2191-2197.

[17] 王兆雪，武法提，郭子涵，等. 创造性思维的跃迁：基于生成式人工智能的多智能体协同系统构建与应用[J]. 中国电化教育，2025(12)：1-10.

[18] 邢凯静. 智能技术辅助下的教师教学效能现状与优化路径研究[D]. 兰州：西北师范大学，2025.

[19] 张山，吴瑛. 认知负荷理论在医学教育领域中的应用进展[J]. 中华护理教育，2023，20(2)：240-244.

[20] 刘祥玲，赵健. 人机协同同伴互助育人模型构建[J]. 开放教育研究，2026，32(2)：89-97.