生物化学与分子生物学课程轻量化知识图谱及智能问答系统构建

张韶涵[1],王晓丽[2],郑 雨[1],胡贵平[1],李 呈[1]
1.北京航空航天大学医学科学与工程学院,北京 100191;2.北京大学公共卫生学院,北京 100191

【摘要】生物化学与分子生物学作为核心基础课,传统的线性教学模式难以解决其概念抽象、知识体系庞大、跨学科整合能力不足等问题,无法满足复合型人才的培养需求。本研究以人民卫生出版社《生物化学与分子生物学》为数据源,构建“轻量化知识图谱+Skill注册式大语言模型(large language model,LLM)智能问答”双层联动系统。整合教材内容为“章节-小节-知识点”三级架构,借助D3.js等前端技术实现了轻量级图谱力导向布局的可视化。系统创新性地设计了“路由-检索-拼装-应答”管线,通过三级降级路由识别用户意图,依据Skill注册表拼装提示词并调用DeepSeek生成回答,实现了知识检索和智能问答功能。与本地知识图谱、纯大语言模型相比,本系统在意图识别准确率、实体检索准确率、事实正确率、回答完整性等指标上均具备最优表现,仅存在极低幻觉风险。系统支持网页端轻量化图谱可视化、学习进度记录与智能问答交互,适配师生不同学习场景,可为生物化学与分子生物学课程数字化教学改革提供实践工具。

【关键词】生物化学与分子生物学;语义知识图谱;大语言模型;Skill路由;智能体
【中图分类号】R-4;G642 【文献标志码】A 【文章编号】1004-5287(2026)05-0588-06
【DOI】10.13566/j.cnki.cmet.cn61-1317/g4.202605002

Construction of lightweight knowledge graph and intelligent question answering system for Biochemistry and Molecular Biology course

ZHANG Shaohan[1], *WANG Xiaoli*[2], *ZHENG Yu*[1], *HU Guiping*[1], *LI Cheng*[1]
1. *School of Medical Science and Engineering, Beihang University, Beijing* 100191, *China*;
2. *School of Public Health, Peking University, Beijing* 100191, *China*

【**Abstract**】As a core foundational course, Biochemistry and Molecular Biology involves such problems as abstract concepts, a vast knowledge system, and insufficient interdisciplinary integration. The traditional linear teaching mode can hardly address these problems or meet the demand for cultivating interdisciplinary talents. This study takes *Biochemistry and Molecular Biology*, published by the People's Medical Publishing House, as the data source and constructs a dual-layer linked system of “lightweight knowledge graph + Skill-registered large language model (LLM)-based intelligent question answering”. The textbook content was integrated into a three-level architecture of “chapter-section-knowledge point”, and front-end technologies such as D3.js were used to visualize the lightweight knowledge graph with a force-directed layout. The system innovatively designed a “routing-retrieval-assembly-response” pipeline. It identifies user intent through three-level fallback routing, assembles prompts according to the Skill registry, and calls DeepSeek to generate responses, thereby achieving knowledge retrieval and intelligent question answering. Compared with the local knowledge graph and the pure LLM, this system demonstrates optimal performance across metrics such as intent

基金项目:国家自然科学基金面上项目(82574066);北京市自然科学基金资助项目(L2608088);北京航空航天大学校级教改项目(403809;414508)
收稿日期:2026-05-29
作者简介:张韶涵,硕士在读,研究方向为医工交叉。
通信作者:李呈,讲师,博士生导师,研究方向为医工交叉。E-mail:li_cheng@buaa.edu.cn

recognition accuracy, entity retrieval accuracy, factual accuracy, and answer completeness, with only a minimal risk of hallucination. The system supports web-based visualization of lightweight knowledge graphs, learning progress tracking, and intelligent question-answering interaction. It is adaptable to different learning scenarios for teachers and students and can provide a practical tool for the digital teaching reform of the Biochemistry and Molecular Biology course.

【Keywords】 Biochemistry and Molecular Biology; semantic knowledge graph; large language model; Skill routing; Agent

生物化学与分子生物学作为核心基础课程，涉及生命体的分子组成、代谢过程、基因调控等生物学原理，是连接生物学与临床医学的桥梁学科[1-3]。在医工交叉融合的背景下，工程背景的学生通常对该课程中抽象的原理感到陌生，缺乏系统的理解，影响了后续医学和工程相关课程的学习效果[4-5]。因此，该课程的教学改革不仅要关注知识内容的有效传递，还应帮助学生建立跨学科的知识框架，提升其系统化理解和应用能力[2]。然而，传统的教学方式往往以课堂讲授为主，难以解决学生知识碎片化的问题。如何提升工程背景学生的学习效果，成为生物化学与分子生物学课程教学改革的关键任务之一。

近些年来，随着大型语言模型(large language model，LLM)技术的突破性发展，教学趋于智能化、数字化。基于LLM的智能问答智能体(Agent)充当学习伙伴，能够处理开放性、非预设的问题，并进行多轮深度交流[6]。但单纯依赖于LLM的智能问答Agent存在局限性，可能会出现回答不准确、逻辑不一致或传播误导性信息等问题，且难以及时融入最新的研究成果和教学要求[7]。知识图谱技术能够通过可视化的方式，系统地展现知识点及其相互关系，帮助学生理解课程内容，避免知识碎片化的问题，但通常需要用户下载安装包、运行安装程序、配置运行环境[8]。网页版知识图谱具有不依赖专用软件的优势，能够在任何设备上通过浏览器访问，学习场景更加灵活、便捷，为学生提供更好的学习支持[9-10]。但知识图谱系统通常需要用户学习特定的查询语言或操作方式，交互不够直观自然，且难以处理需要综合理解与创造性表达的复杂问题。综上所述，单纯依赖知识图谱难以处理开放式提问，单纯依赖LLM又难以保证内容的准确性。因此，本研究以课程教材为核心数据源构建语义知识图谱，融合LLM作为可调用的扩展能力按需补全语言生成，并通过预设Skill注册表实现“意图-回答模板”的统一路由，帮助学生更好地理解复杂的生物化学原理与分子过程，为后续医学和工程的深度融合奠定基础。

1　研究方法

1.1　语义知识图谱构建

根据人民卫生出版社的《生物化学与分子生物学》(第9版)教材及对应的课程教学大纲和教材内容，梳理本课程的核心概念和重要知识点。通过在知识点节点中增设aliases(别名/同义词)字段来提升检索的鲁棒性，如“糖酵解”的别名为“葡萄糖分解”或“EMP途径”。利用D3.js实现知识图谱的力导向布局，为知识点建立关系网络。整体系统采用单HTML文件部署方式，可在普通教学终端或学生个人设备上直接访问，满足轻量化教学应用需求。

1.2　语义知识图谱系统架构

系统整体划分为四层设计，其中数据层负责知识图谱与学习记录存储，知识组织层负责将图谱检索结果转换为LLM可理解的上下文，Agent核心层负责意图识别、知识检索、Skill路由与Prompt拼装，交互显示层负责图谱可视化和智能问答结果呈现。

1.2.1　数据层

数据层主要负责存储系统的核心知识内容，包括知识图谱数据库和学习记录数据库。知识图谱数据库以教材为核心数据源，遵循统一的知识纳入标准搭建知识体系。知识纳入标准遵循教材核心性的原则，依据知识点的重要性区分，重要性＞6分的概念重点纳入，重要性≤6分的为补充性知识点；概念描述必须包含完整的定义或解释，且每个概念必须归属于某个小节。每个知识点包含名称、概念、aliases、关系、重要性等级和所属章节等信息，每条关系包含关系类型、关系描述和关系强度等信息。aliases生成规则包括以下4条。①同义表述：教材中明确提出“简称”“又称”等词，视其为别名；②概念同一性：教材内表述不同但内涵完全等价的概念，可等价映射，将其互设为别名；③缩写与全称：若概念包含英文全称，则自动生成其英文缩写作为别名；④专家增补：以上规则无法覆盖的同义俗称，由领域专家手工增补。重要性等级判定如下：1～4分为极其具体的细节或边缘性知识点；5～6分为支撑性概念、基础定义或历史背景，属于重要的铺垫性概念；7～8分为支撑关键概念的

机制、分类或重要研究方法等；9～10 分为理解全篇的“第一性原理”。关系强度判定遵循以下原则：为每种类型预设基础强度值，包含关系根据层级进行微调，核心逻辑链强度高于间接关联。专家审核流程分为三步：首先，通过大模型，根据目录和教材内容生成初始节点与关系；其次，两位领域教授独立对所有节点按上述维度打分；最后，若两人分差超过 2 分，需在专家会议上说明理由，并讨论修正，达成共识后修改数据。学习记录数据库存储本地用户学习进度，记录已学习的知识点。

1.2.2 知识组织层

知识组织层负责将原始知识图谱信息转化为 LLM 可理解的语句，并进行结构化输入。本层将模板库、知识嵌入模块和上下文管理模块的功能进行整合。构建 Prompt（提示词）三段式结构，按照 System（系统角色与任务设定）、Context（知识图谱检索结果、用户学习进度等上下文信息）、User（用户当前问题）三个部分拼接提示词。

1.2.3 Agent 核心层

Agent 核心层控制知识图谱的智能问答系统，负责分析问题类别、检索相关知识、构建提示词，并调用 LLM 进行回答。构建智能问答管线（Step A～Step E）如图 1 所示。Step A 采用三级降级路由，根据用户所提问题中的关键词进行匹配，若符合强规则，则直接确定 Skill 类型（置信度≥0.85），Skill 注册表如表 1 所示。若不符合强规则，则进入软规则进行关键词匹配，并计算语义相似度，取匹配度得分最高的 Skill 类型，匹配度得分≥0.6 则采纳，匹配度＜0.60 则将原始问题连同最近 1～2 轮对话历史发送至 LLM，要求其输出意图分类和 Skill 类型，返回结果的置信度须≥0.75。Step B 则从问题中抽取标准化的知识点名称，并结合多轮对话历史进行指代消解。Step C 依据抽提的实体，在知识图谱中进行精准匹配（置信度≥0.95）、包含检测（含 aliases，置信度≥0.85）和描述匹配，匹配结果依据匹配度和关系强度进行排序，取 Top-K 候选（K=5，基于知识图谱平均关系数 0.78 和验证集召回率实验确定，确保 80% 以上问题的正确答案落入候选集）。Step D 根据 Step A 阶段确定的 Skill 类型，将检索到的知识点套入模板，形成结构化上下文，最终拼装为三段式 Prompt（提示词）模板。Step E 将拼装好的 Prompt 发送至远端 LLM 模型，参数设置为 max_tokens=1 200、temperature=0.7、top_p=0.95、frequency_penalty=0.2、presence_penalty=0.1、stream=false，进一步补充临床或工程应用等扩展知识，生成最终回答。当 API 请求失败或超时，Step E 自动跳过 LLM 调用，转由 Skill 自带的兜底渲染模板将图谱事实直接结构化输出。

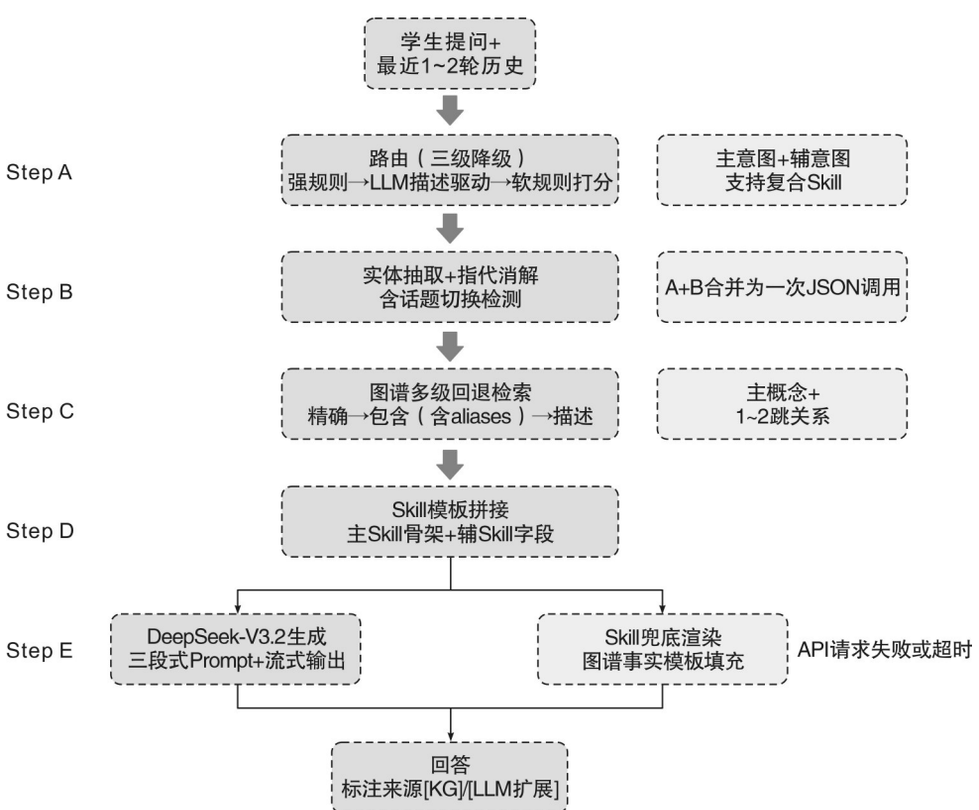


图 1 智能问答管线

表 1　Skill 注册表

Skill 名称	描述（用于 LLM 路由提示）	强规则触发关键词	检索方式	在线 Prompt 模板要点	离线兜底渲染
concept_definition	回答“…是什么/…的定义”等概念定义类问题	“什么是…”“…是什么”“…指什么”“…的定义”	单实体精确匹配（含 aliases）＋1 跳关联	定义＋重要性＋章节＋关联节点（KG）＋应用扩展（LLM）	四段式输出节点字段
concept_compare	比较两个或多个概念的异同	“…和…的区别”“对比…”“…与…的异同”	多实体并列检索＋关系交集	维度化对比表＋共同点/差异	双节点并列渲染
process_explanation	解释生化机制、流程与时序步骤	“…的过程”“…的机制”“…的步骤”	主实体＋时序关系遍历＋关键酶节点	时序分解＋核心调控位点＋关键参数	按关系顺序输出
relation_analysis	分析因果、相互作用与影响类问题	“为什么…”“…如何影响…”“…和…的关系”	主实体 1～2 跳关系遍历＋关系强度排序	因果链＋关键通路＋反馈调控	关系列表＋强度排序
open_qa	兜底：复合问题、规则未命中场景	规则未命中＋LLM 路由置信度低＋API 不可用	实体软匹配＋Top-K 弱相关	自由回答；无对应知识点注明教材未收录	通用知识卡片＋学习建议

1.2.4　交互显示层

交互显示层为满足用户的学习需求，提供知识图谱可视化和智能问答 Agent 两种交互显示。知识图谱可视化包括主视图区和实体详情面板。主视图区包含节点、边、图例及缩放控制按钮等元素，可根据节点的颜色和大小判断其所属层级和是否进行学习。实体详情面板由章节列表面板、详细信息面板和关系管理面板组成，显示当前知识点所属层级、概念、重要性和相关关系等。智能问答 Agent 包含消息展示区、输入框和 API 配置按钮，输入问题后系统通过智能问答管线进行处理，根据需要调用 LLM 进行回答并展示答案。

1.3　系统回答质量评估

本研究从生物化学与分子生物学课程知识图谱中，随机抽取 50 个问题形成标准化测试问题集，问题集覆盖五大问题类别，评估采用专家人工评分的方式。各维度评分如下：①意图识别准确率，1 分为不准确，2 分为准确；②实体检索准确率，1 分为不准确，2 分为准确；③事实正确率，1 分为不准确，2 分为准确；④回答完整性，1 分为回答缺失，2 分为回答较为完整，3 分为回答完整；⑤幻觉率，1 分为严重幻觉，2 分为轻微幻觉，3 分为无幻觉。

2　结果与分析

2.1　语义知识图谱的可视化

本研究构建的课程语义知识图谱，依托 Web 前端技术，将教材核心知识点进行层级化和结构化的可视化呈现。知识图谱层级网络的节点数量为 1 193 个，其中章节、小节和知识点节点数依次为 16、73 和 1 104 个，关系数量为 1 234 个，主要包括结构与组成关系、因果与决定关系、过程与演变关系、功能与用途关系、调控与影响关系、识别与结合关系、状态与比较关系、来源与去向关系八大关系类型。每个节点均为独立的知识实体，根据其层级的不同被赋予不同的颜色和大小，其中最大的蓝色节点为一级节点（章节），绿色的节点为二级节点（小节），最小的红色节点为三级节点（知识点）。

针对教学者和学习者，根据其需求设计语义知识图谱的核心功能。面向教学者设计了整体预览展示功能和实体详情面板，整体预览帮助快速建立对课程的宏观认知，实体详情面板展示单个知识点名称、所属章节、概念及关联知识点等信息，帮助梳理不同章节知识点间的内在关联，将抽象的知识可视化。针对学习者用户的需求，增加学习进度功能，根据自身学习节奏对知识点进行深度学习，完成学习后该知识点将自动被标记为黄色，并在进度面板记录已学习知识点数目，点击页面上方的“学习进度”按钮，展现学习者对教材知识的掌握情况，以“打卡”的形式完成整本教材的学习，提升学习兴趣。

2.2　语义知识图谱的检索功能

用户可通过知识点名称、所属章节、所属小节等单一或组合条件对知识点进行检索。例如，在核心概念浏览页面输入“活性中心”可直接定位该知识点的所属章节及关联内容。该过程依托多级回退检索算法：系统先在 aliases 字段中命中“酶活性中心”，随后自动展开其所属章节“酶的结构与功能”及关联的“底物结合”“催化机制”等关系节点，能够快速、准确地获取该知识点的全部信息，解决了传统教材检索中需要逐页查找、效率低下的问题。

2.3　语义知识图谱智能问答管线效果

本研究构建的语义知识图谱内置智能问答 Agent，支持本地知识图谱事实和 LLM 增强两种回答来源，选取概念定义、复合意图和离线兜底 3 个典型场景展示系统效果。

2.3.1　问题类型识别与语义分析

用户在智能问答模块进行提问，智能问答 Agent

首先对问题进行类型识别和关键词提取，根据问题类型建立不同模板进行回答。根据本课程的特点和学生易混淆的问题，研究定义 5 种问题类型，分别为概念定义类、概念比较类、过程解释类、关系分析类和一般问答类，根据问题中关键词识别分类，例如问题中包含“什么是……?”或“……是什么?”为概念定义类问题，问题中包含“过程”或“机制”为过程解释类问题。

2.3.2 知识检索

通过混合式检索方法，从知识图谱中检索与用户问题相关的概念和关系。当用户的问题中明确提及某一概念名称时，系统会通过精准匹配以及概念、描述的包含匹配来检索相关概念。对于关系分析类问题，首先检索查询词概念，再检索、筛选该词的关系列表。对于概念比较类问题，需同时检索 2 个概念的完整信息及概念相关关系，通过模型进行对比总结后回答。

2.3.3 本地知识图谱模式

用户在本地知识图谱模式进行提问，智能 Agent 可直接调取语义知识图谱内知识点数据给出答案。

2.3.4 LLM 增强模式

用户可输入自己的 API 密钥切换到大语言模型增强模式，在该模式下，智能 Agent 将用户所提问题与当前用户正在查看的章节、概念、已学习知识点及总体学习进度整合为系统提示词的一部分，发送至 LLM。以 DeepSeek-V3.2 大模型为例，将本地知识不足以回答的复杂问题通过模型进行知识补充，在进行多轮连续问答时，会保留多轮历史对话作为上下文。

2.4 系统问答质量评测与对比分析

本研究随机抽取 50 个问题形成标准化测试问题集，其中，章节、小节和知识点类问题分别为 2、8 和 40 个。从意图识别准确率、实体检索准确率、事实正确率、回答完整性和幻觉率 5 个维度，对本地知识图谱、纯 LLM 和 LLM 增强 3 种回答模式进行对比。

本地知识图谱模式和 LLM 增强模式的意图识别准确率分别为 92% 和 98%，实体检索准确率分别为 100% 和 98%，LLM 增强模式仅出现 1 次未被检索到，表明提问可在知识图谱中被相对准确地检索。在事实正确率方面，LLM 增强模式以 100% 的正确率优于本地知识图谱模式（88%）和纯 LLM 模式（90%），表明依托问答管线后，LLM 增强模式在回答正确性方面更具优势。在回答完整性方面，LLM 增强模式为 94%，相较于本地知识图谱模式（40%）和纯 LLM 模式（50%），其回答完整性优势明显。在幻觉率方面，纯 LLM 模式和 LLM 增强模式的幻觉率分别为 40% 和 8%，LLM 增强模式依托问答管线使幻觉率显著降低。

综上所述，LLM 增强模式在 5 个维度中表现良好，优于本地知识图谱模式和纯 LLM 模式。

3 讨论

在大数据、人工智能和知识图谱快速发展的背景下，高等教育正在经历数字化转型。联合国教科文组织《2024 年全球教育监测报告》显示，目前全球已有 78% 的发达国家把人工智能纳入其教育现代化战略[11]。在此背景下，生物化学与分子生物学从传统理论课程转变为跨学科、重整合的核心课程，借助知识图谱技术对课程内容进行系统化梳理和整合，优化课程体系、创新教学模式、支持个性化学习，推动课程教学改革。

传统语义知识图谱主要依赖于可视化和关系查询功能，虽提供了清晰的知识框架，但仅能学习教材中的固有知识点，无法解决学习中遇到的问题和难点。因此，本研究在传统知识图谱的基础上，内置智能 Agent 问答系统，帮助用户解决学习中遇到的问题和难点。同时，生物化学与分子生物学是一个持续发展的学科，传统教学模式往往面临教材更新滞后的问题，本系统增设了 LLM 用以补充教材中知识点的不足。此外，知识图谱的构建能够实现对学科内容的动态更新[12]。在本研究构建的知识图谱中，随着新知识点的增加，相关节点和关系持续更新，辅助学生了解最新的科研成果和学科发展趋势。未来的优化重点期望让图谱覆盖得更全、关系更准确，通过 Skill 注册表的细化与回答的来源标注约束 LLM 的输出严格落在学科边界之内。

此外，知识图谱的准确性直接影响到其在教学中的应用效果，若节点与边的关系设置不当或存在遗漏，将导致知识图谱无法准确地反映学科的真实结构和知识网络[13]。在医工融合的背景下，生物化学与分子生物学课程不仅涉及生物学知识，还要与医学、工程学、计算机科学等多个学科的知识交织[12,14]。跨学科的融合往往面临知识体系差异、概念表述不一致等问题[15]。本系统在 LLM 增强模式下的回答质量优于本地知识图谱模式，但仍存在一定局限性。例如，用户提问口语化、表述不完整和使用错别字等，LLM 在进行回答时易出现偏差。此外，用户提出一些推理过于复杂的问题时，LLM 虽语言理解能力出色，但严谨性方面仍显不足。因此，如何在知识图谱中有效地整

合不同学科的知识，并确保这些跨学科概念的准确转换与衔接，提升自然语言处理系统在复杂学科背景下的理解能力，是未来优化知识图谱应用效果的重要方向之一。

4　总结与展望

本研究构建了面向生物化学与分子生物学课程的轻量化语义知识图谱，并在此基础上设计了“路由-检索-拼装-应答”的智能问答管线。系统以本地知识图谱保证事实的准确性，以 Skill 注册表组织不同学习需求下的回答策略，以 LLM 提供语言扩展能力。该系统兼顾教学者和学习者双重需求，有助于提升学生对知识的理解、提高学习兴趣，并为生物化学与分子生物学课程教学提供技术支撑。未来研究将聚焦于四个方面：一是通过“教材分块-LLM 抽取-专家校验”提升图谱构建效率；二是扩展实验操作、考点速记和临床案例等场景化 Skill；三是引入语义向量检索以提升口语化提问理解能力；四是构建真实学生提问评测集，对比纯 LLM、本地知识图谱和本研究管线在准确性、完整性与教学适用性上的差异。在此基础上，进一步强化知识图谱的动态更新能力，为培养高层次医工融合人才提供更有力的支撑。

参考文献

[1] 马灵筠，席守民，杨五彪. 生物化学与分子生物学在医学教育中的重要性及教学探索[J]. 西北医学教育，2010，18(2)：319-321.

[2] 靳祎，王怀颖，邸科前，等. 新医科视域下临床医学专业生物化学与分子生物学课程的学科融合式教学改革[J]. 医学研究与教育，2025，42(3)：63-73.

[3] 孙丽郡，孟繁军，徐行健. 课程知识图谱构建技术研究综述[J]. 计算机工程，2025，51(11)：1-21.

[4] 刘金文，文晓岚，莫智雯，等. 医学类专业生物化学与分子生物学多元化教学改革探索[J]. 教育观察，2024，13(4)：82-84.

[5] 朱华，陈莉，陶毅明，等. 基于建构主义理念的教学方法在医学生物化学课程中的探索与实践[J]. 中国高等医学教育，2023(7)：74-76.

[6] 卢宇，余京蕾，陈鹏鹤. 基于大模型的教学智能体构建与应用研究[J]. 中国电化教育，2024(7)：99-108.

[7] 王佑镁，王欣颖，柳晨晨. 教育领域生成式人工智能应用的伦理风险管理框架研究[J]. 电化教育研究，2024，45(10)：28-34.

[8] 吕银云，朱亚先. 浅谈虚拟教研室中知识图谱的搭建与应用[J]. 大学化学，2023，38(10)：49-55.

[9] 范佳荣，钟绍春. 学科知识图谱研究：由知识学习走向思维发展[J]. 电化教育研究，2022，43(1)：32-38.

[10] 周东岱，董晓晓，顾恒年. 教育领域知识图谱研究新趋向：学科教学图谱[J]. 电化教育研究，2024，45(2)：91-97.

[11] UNESCO. Global education monitoring report 2024/5, leadership in education: Lead for learning[R]. Paris: UNESCO, 2024.

[12] 郑君芳，陈子钰，王雯. 疾病表型-分子机制双轴融合的智慧教学体系构建[J]. 医学教育管理，2025，11(5)：523-528.

[13] Chen J, Dong H, Hastings J, et al. Knowledge graphs for the life sciences: Recent developments, challenges and opportunities[J]. Transactions on Graph Data and Knowledge, 2023, 1(1): 5.

[14] Douard N, Cavallucci D, Samet A, et al. Large-scale transformer-based topic graphs identify thematic links between engineering and biology[J]. Scientific Reports, 2025, 15: 29256.

[15] 张瑜，李拓宇，叶民. 整合与重塑：新时代交叉科学研究演进与展望[J]. 工程研究：跨学科视野中的工程，2026，18(4)：559-574.