科学技术与教育 2026年第9期
ISSN: 3079-4455

多智能体协作的个性化教学资源生成与评估系统

李湘[1,*]，郭欣茹[1]，田莹[1]，梁紫函[1]，尹克心[1]，郝振[1,2]，刘婉月[1,2]
[1]河南工业大学人工智能与大数据学院，河南郑州，中国
[2]科大讯飞股份有限公司，安徽合肥，中国
*通讯作者

【摘要】当前智慧教育领域多智能体学习平台普遍存在教学内容同质化、学情量化维度单一、人工智能生成内容可靠性低以及教学管理流程繁琐等问题，难以真正实现因材施教。现有产品尚未打通“学习情况诊断—资源生成—内容审核—练习反馈—动态评价”完整教学闭环，多模态教学资源生成、学情评估、内容校验模块相互割裂，制约个性化教学落地效果。本文设计并实现 LearnWeave 多智能体个性化知识生成系统，构建基于 LangGraph 的事件驱动三智能体协同调度架构，融合 RAG 检索增强、混合向量检索和多模态自动渲染技术，实现讲义、代码案例、思维导图、教学音频和视频五类材料的一体化生成。系统通过学习情况诊断、领域知识生成和审核判定三个智能体实现串行核验，采用 SQLite 与 ChromaDB 分离存储，构建六维量化评价模型，并区分学生、教师和管理员权限，形成完整教学闭环。测试表明：系统 AI 生成教学内容幻觉率低于 3%，知识点匹配准确率较通用大模型直出方案提升 15%－21%；课程材料检索延迟低于 180ms，多模态材料生成效率提升超过 60%，可完整支撑高校计算机类课程分层个性化教学全流程。

【关键词】多智能体协作；检索增强生成；多模态教学资源生成；智能学习情况评估；分层自适应学习

1.引言

数字化智慧教育持续普及，高校广泛采用线上学习平台辅助线下授课，但多数在线学习平台采用标准化课件和固定学习路径，无法适应学生在知识基础、编程实践能力和自主学习节奏方面的差异。通用大模型还存在内容脱离教材、生成失真、难度不匹配和缺乏专业审核等问题，无法直接用于课堂教学。教师难以精准掌握全班薄弱点，备课、编题和实验代码评分负担沉重，平台管理员也缺少统一运维工具。Wang 等[1]的高校研究显示，个性化学习环境的有效性取决于学习者建模、资源适配和教师参与的协同。

具体行业痛点体现在四方面：一是多模态教学素材联动生成能力薄弱，文本、代码、音视频知识点难以实现时序与语义对齐；二是学情评估维度单一，仅依靠答题分数无法综合刻画学生综合学习能力；三是大模型知识生成易出现幻觉，缺少知识库溯源与多层级内容校验机制；四是系统角色功能耦合严重，学生自主学习、教师教学管控、管理员运维缺乏独立标准化流程。相关系统综述也指出，学习分析、生成式模型与自适应推荐的整合仍是高等教育个性化学习的关键研究方向[2]。

Och 等[3]最早提出多对一的多智能体协同处理思路，将多源信息共同输入以决定输出结果。Zoph 等[4]将多对一协同理论应用于神经网络，实现了多编码器单解码器架构。Garmash 等[5]提出多模型融合方法，将不同模型的输出结果进行融合以确定最终结果。Dabre 等[6]提出的多源数据拼接方法，无需改变模型结构即可实现多源信息输入。Libovicky 等人[7]在 Transformer 模型架构之上探究了多种编码器融合方法。现有智能教育平台主要通过改进生成模型、融合多源数据、优化评估体系三类路径实现个性化教学，但仍存在诸多局限：部分平台仅实现基础功能无 AI 个性化内容生成能力；部分简易 AI 答疑工具不具备学习者画像量化、多模态资源自动生成、三角色权限分离管理体系；主流自适应学习平台仅由单一智能体完成内容生成，缺少独立裁判校验环节。近期多智能体研究进一步表明，基于协商与共识的多模型协作能够提升复杂推理的稳定性[8,9]。

针对上述问题，本文研发 LearnWeave 多智能体个性化知识生成系统，搭建基于 LangGraph 的事件驱动三智能体协同调度架

构，融合 RAG 检索增强、混合向量检索、多模态自动渲染技术，完成文本讲义、代码案例、思维导图、教学音频、视频五类素材一体化生成；系统分层实现学情诊断 Agent、领域知识生成 Agent、审核裁判 Agent 串行流转校验，依托 SQLite 业务库与 ChromaDB 向量库分离存储教学数据，构建六维能力量化 AI 智能评估模型，区分学生、教师、管理员三类权限角色形成完整教学闭环。

本文的贡献如下：1）提出三智能体串行协作流水线，实现学情诊断—领域生成—审核裁判全流程自动化流转与可视化追溯；2）构建课程专属混合 RAG 检索架构，AI 输出强制绑定教材来源，有效抑制大模型幻觉；3）设计六维学习者画像量化模型，区分三类学习分层路径，动态调整全部学习资源难度；4）实现“一源多生”全自动多模态生成引擎，可基于单一知识点同步生成五类教学资源；5）提出双数据库分层存储设计，支持私有化部署保障教学数据安全。

2.相关工作

多智能体协同技术起源于自然语言处理领域。Och 等[3]提出了多对一的多语言翻译思想，实现了多对一多语言统计机器翻译模型，但统计机器翻译模型不能进行端到端的训练。Zoph 等[4]将多对一多语言翻译的思想运用到神经网络中，编码器和解码器采用 LSTM 实现端到端的训练，多个编码器的结果通过模拟 LSTM 进行融合。Garmash 等[5]提出多模型融合方法，即单一语言对训练翻译模型，将不同语言对的翻译模型进行融合，该方法需要训练多个子模型，实验耗时较长。Dabre 等[6]提出多种源语言句子拼接的方法，重建数据集，不需改变模型结构实现多源输入。这些多智能体协同思想为教育领域的多 Agent 协作提供了理论基础。

2016 年，Ha 等[10]提出了一种统一的方法，在不改变模型结构的前提下，无缝地将原始神经机器翻译扩展为多语言翻译。Johnson 等[11]在 2017 年采用了类似的方法，在源端每一句开头都加标记，以表示相应的目标语言。这种语言标记法的思想被本系统借鉴用于区分不同学习者画像和课程类型。近年来，Blackwood 等[12]提出共享除注意力机制之外的所有参数，以降低多任务模型的训练成本；Sachan 等[13]研究了 Transformer 模型中各种参数的共享策略；Platanios 等[14]建议共享整个网络，同时使用上下文参数生成器。但上述研究多聚焦于模型层面的优化，未结合教学业务流程形成完整的学习闭环。

在智能教育领域，国外多智能体教育系统多基于海外通用大模型构建，侧重自适应刷题功能，缺少面向国内高校计算机、大数据专业课程的本地知识库检索能力，无课程实验代码自动批改模块；主流自适应学习平台仅单一智能体完成内容生成，缺少独立裁判校验环节。国内现有智慧教学平台分为两类：一类仅实现直播、课件上传、在线考试基础功能，无 AI 个性化内容生成能力；二类简易 AI 答疑工具，不具备学习者画像量化、多模态教学资源自动生成、三角色权限分离管理体系。现有产品均未打通“学情诊断—资源生成—内容审核—练习反馈—动态评估”完整学习闭环，多智能体协同调度、本地 RAG 知识库、六维能力画像量化三大技术点存在明显空白，为本系统研发提供了创新空间。在知识增强方面，RAG 研究已从基础检索扩展到答案编码、信息精炼和自动化评估[15-17]，为课程知识库溯源和生成质量控制提供了方法基础。

3.课程知识库构建与数据存储

本系统以高校计算机类 Python 数据分析课程为核心应用场景，人工构建课程专属知识库，利用课程讲义、实验手册、习题集作为原始数据，通过文本解析、分块、向量化、索引构建等步骤，形成支持混合检索的多模态课程知识库。知识库的构建过程分为 4 个部分：原始教学资料的收集与整理、文本解析与分块处理、向量化与索引构建、质量校验与人工筛选。

原始教学资料的收集与整理：原始资料分别是课程讲义、实验指导手册、课后习题集、代码案例库、数据集说明文档。所有原始资料统一转换为纯文本格式，去除格式标记和冗余内容。文本解析与分块处理：文本分块采用教学场景专属策略，按讲次、知识点边界切割长文档，控制单块文本 token 长度在 200-500 之间，兼顾上下文完整性与检索精度。每个文本块自动挂载元数据，包括课程名称、讲次编号、知识点标签、难度等级、来源文档 ID 和段落位置。

向量化与索引构建：分块完成后，调用 sentence-transformers 嵌入模型将课程文本转换为高维向量，存入 ChromaDB 向量数据库。同时构建 BM25 关键词索引，将文本块

的关键词、词频信息存储在 SQLite 业务数据库中。双索引架构使得系统能够同时支持向量相似度检索和关键词精确匹配，两种检索结果通过加权融合算法进行排序。质量校验与人工筛选：最终得到的课程知识库包含 60 份教学资料，文本块总数约 800 个，覆盖 Python 基础、数据处理、数据分析、可视化、机器学习等全部课程知识点。

系统采用业务数据与向量知识数据分离的双数据库设计，其分层存储架构如图 1 所示。SQLite 作为业务数据库，存储用户与角色数据、学习者画像数据、课程与讲次数据、实验和数据集元数据、学习行为与每日任务、笔记、对话历史、答题记录及评估数据。ChromaDB 作为向量数据库，存储课程讲义、实验手册、题目等文本的向量化索引。双数据库通过知识点 ID 与文档 ID 进行关联，既保证了业务数据的事务性与读写稳定性，又实现了知识内容的高效语义检索，同时便于私有化部署时的数据隔离与安全管控。

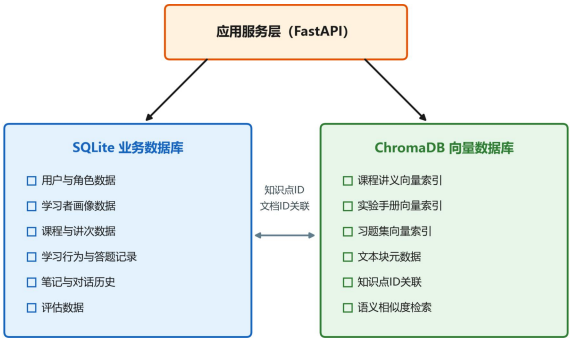


图 1.双数据库分层存储架构

4.系统总体架构与核心技术

本论文从智能体协同机制、检索增强技术、多模态内容生成、学情评估模型四个方面实现 LearnWeave 个性化知识生成系统。智能体协同机制通过三 Agent 串行流转实现内容生成的全流程管控；检索增强技术通过混合检索与溯源校验抑制大模型幻觉；多模态内容生成实现“一源多生”的五类资源自动产出；学情评估模型构建六维量化体系支撑分层教学。

系统采用前后端分离架构，整体由浏览器前端、FastAPI 服务层、Agent 编排层、知识检索层和外部能力层五部分组成，系统总体组成如图 2 所示。前端使用 Next.js、React 和 TypeScript 实现页面展示与用户交互，后端使用 FastAPI 和 Pydantic 提供认证、画像、知识检索、资源、题库、评估及教学管理接口。智能体协作层包含三个职责不同的核心 Agent，知识检索层使用课程知识库并根据部署状态选择向量索引或关键词检索，外部能力层按配置连接大语言模型服务。

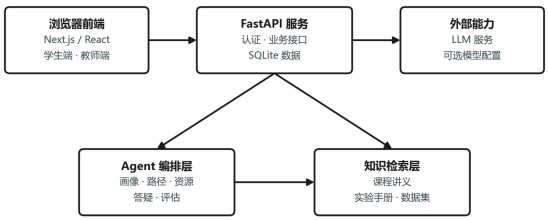


图 2.系统总体组成

系统功能模块围绕学生自主学习、教师教学管理、管理员系统运维三大功能集群组织，认证与权限模块位于业务模块之前，学生和教师共享课程资源、数据存储及后端服务，但根据账号角色进入不同功能范围，功能模块结构如图 3 所示。

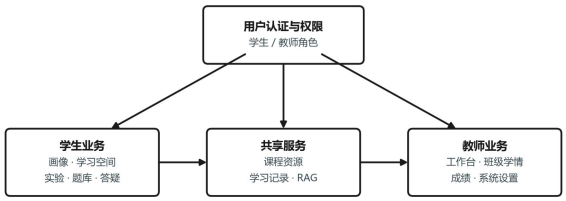


图 3.系统功能模块结构

4.1 三智能体串行协作机制

系统采用固定串行工作流：学习者画像输入→学情诊断 Agent 分析知识盲区与推荐难度→知识库检索匹配对应素材→领域知识生成 Agent 生成五类多模态资源→审核裁判 Agent 多维度校验→校验通过资源入库，校验失败自动回流生成节点重新优化内容，形成闭环自修正机制。完整的个性化知识生成与学习闭环流程如图 4 所示。

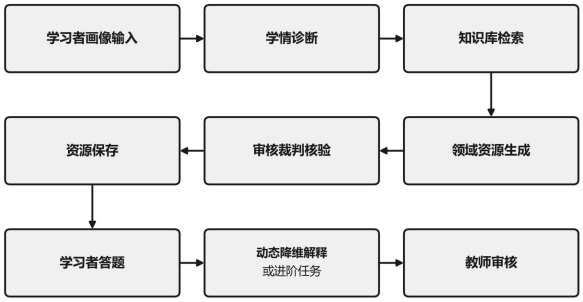


图 4.个性化知识生成与学习闭环

智能体协作层包含三个职责不同的核心 Agent，其串行协作流程如图 5 所示：学情诊断 Agent 负责分析学习背景、知识基础、技能盲点、学习目标和推荐难度，通过多轮对话逐步收集学习者信息，综合形成六维能力画像和知识盲区清单；领域知识生成 Agent

负责检索垂直领域知识库并生成知识讲解、实操指南和分阶测试题，根据学情诊断结果动态调整内容深度和难度，同时调度五类资源生成子任务；审核裁判 Agent 负责核验专业性、知识覆盖、难度匹配、逻辑一致性和引用证据，对生成内容进行多维度质量打分，不符合要求的内容自动退回生成节点重新优化。

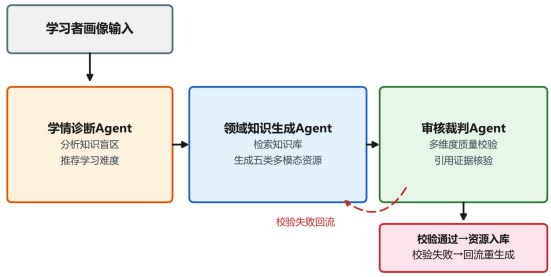


图 5.三智能体串行协作流程

基于 LangGraph 有向任务图搭建智能体流水线，每个 Agent 为独立可重启节点，支持任务暂停、重置、重试；节点间标准化传递画像、知识库检索结果、生成内容、审核结论结构化数据，实现任务解耦。系统根据基础薄弱、标准进度和高阶三类标签，动态调整讲义深度、代码注释、视频时长和题目难度。

4.2 核心技术实现

4.2.1 混合 RAG 防幻觉机制

针对通用大模型脱离教材生成内容、易出现幻觉的问题，系统构建课程专属混合 RAG 检索架构，其技术架构如图 6 所示。融合 ChromaDB 向量相似度检索与 BM25 关键词检索，双重匹配校内课程原始文档，优先使用教材内容作为 AI 生成上下文，降低大模型凭空编造知识点概率。向量检索能够捕捉语义相似性，召回含义相关但表述不同的文本块；关键词检索能够精确匹配专业术语和特定概念，两种检索结果通过加权融合算法进行排序。

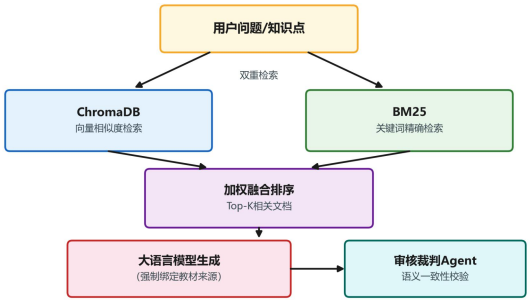


图 6.混合 RAG 检索增强防幻觉架构

每一段 AI 生成内容自动挂载匹配知识库文档 ID、段落位置，页面展示可点击来源；审核裁判 Agent 校验生成内容与原文语义一致性，偏离教材内容直接判定不通过，拒绝入库。内容溯源与引用校验算法确保所有 AI 生成文本都有权威教材来源支撑，从根源降低大模型知识幻觉。

4.2.2 “一源多生”多模态内容生成技术

多模态融合技术统筹文本、代码、思维导图、音频、视频五类教学资源，实现同一知识点下多模态素材同步生成、统一索引、联动推送，构建“一源多生”全自动生成引擎，其生成架构如图 7 所示。仅输入课程知识点描述，系统自动并行生成五类配套学习资源，大幅降低教师备课成本。传统教学模式下，教师准备一个知识点的完整教学材料需要分别撰写讲义、绘制思维导图、编写代码案例、录制讲解音频、制作教学视频，耗时数小时甚至数天；本系统将这一过程自动化，单知识点五类资源的生成时长控制在 3 分钟以内。

文本分析模块完成课程讲义、实验手册、题目全类文本数据的解析、分块、向量化、检索、生成与校验。音频模块通过 TTS 文本转语音、音文对齐算法生成与讲义匹配的教学音频，采用教学专属语音音色库。视频模块依托 Manim 代码动画渲染引擎，自动生成代码运行、数据运算、流程讲解动态画面，结合音频与字幕合成完整教学视频，全链路无人工视频制作。审核裁判 Agent 同步校验同一知识点下所有模态内容的逻辑一致性，保障跨模态内容统一。

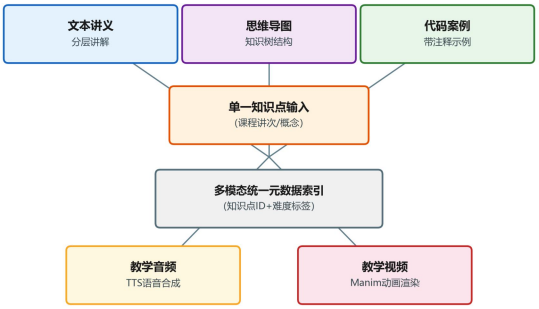


图 7.“一源多生”多模态教学资源生成引擎

4.2.3 六维学情量化评估模型

突破传统系统单一答题正确率的评估方式，构建六维能力加权评分模型，从理论基础、编程能力、实践操作、问题排查、数据思维、自学能力六大维度动态加权计算综合得分，通过雷达图可视化直观展示学生短板。三类学习者画像的六维能力对比如图 8 所示。区别于传统系统仅统计正确率，本模型能够完整刻画学生综合学习能力短板，定

位深层知识盲区。

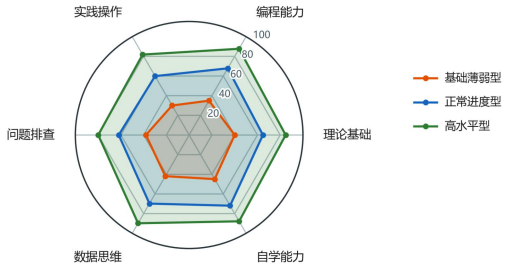


图 8.六维能力量化评估模型

系统结合错题知识点关联挖掘、学习行为时序分析，自动识别知识薄弱点，生成个性化补强学习方案。评估结果与学习资源动态联动，识别薄弱知识点后，自动推送对应讲次讲义、专项习题、实操实验视频，实现评估即补强。教师端可自动汇总全班共性薄弱项，批量生成班级专项补强教学素材，支撑分层教学干预。

4.2.4 安全设计

系统通过登录状态和角色信息限制页面访问，密码以散列形式保存，采用 bcrypt 算法进行密码哈希，浏览器仅保存会话凭证，使用 JWT 令牌进行身份认证。模型服务密钥和数据库连接信息由服务端配置管理，存储在环境变量或配置文件中，不在前端页面展示，也不随 API 响应返回。用户输入内容经过敏感信息拦截，采用正则表达式和关键词匹配识别密码、密钥、身份证号、手机号等敏感信息，避免密码、密钥或个人隐私数据泄露。不同角色账号数据严格隔离，学生、教师、管理员仅可访问权限范围内的功能与数据，后端接口进行角色权限校验，防止越权访问。系统采用业务库与向量库双分层存储、密码哈希加密、服务端敏感配置隔离、角色权限多层拦截安全架构，私有化部署适配 Windows、Linux 本地环境，完全满足高校数据安全要求。此外，系统支持操作日志审计，记录用户登录、资源访问、数据修改等关键操作，便于安全事件追溯。

5.系统测试与实验

本文以高校计算机类 Python 数据分析课程为测试场景，使用系统自建的课程知识库与测试账号，从功能完整性、性能指标、内容质量三个维度开展测试。首先验证系统核心功能的可用性与角色权限隔离效果，期间对比不同学习者画像下的资源差异；其次对检索延迟、生成效率等核心性能指标进行测试；最后对 AI 生成内容的幻觉率、知识点匹配准确率进行量化评估，并与通用大模型直出方案进行对比。

5.1 实验数据

自建的 Python 数据分析课程全套讲义、实验手册、习题集共 60 份教学资料作为实验数据，文本块总数约 800 个。分别从全部语料中随机抽取 1000 条涵盖基础概念、代码实操、问题排查三类知识点的测试问题集，用于评估内容生成质量与检索准确率。同时构建基础薄弱、正常进度、高水平三类标准学习者画像，用于验证分层个性化生成效果。代码采用 Python 语法解析工具进行分词，其它文本采用 jieba 分词工具进行数据预处理。

5.2 实验设置

前端运行环境为 Node.js 20.9.0 及以上，依赖使用 pnpm 管理；后端运行环境为 Python 3.10+，采用 Uvicorn 作为 ASGI 服务器；向量数据库使用 ChromaDB 本地部署，业务数据库使用 SQLite。大语言模型采用兼容 OpenAI 接口的开源模型，嵌入模型使用 sentence-transformers 系列模型。所有测试均在本地私有化部署环境中完成，无外部云端服务依赖。核心技术栈包括 Next.js 14、React 18 、 TypeScript 、 FastAPI 、 LangGraph 、 ChromaDB 等。

5.3 基线模型

本文以三类基线方案作为对比基准：一是通用大模型直出方案，即不使用 RAG 检索增强，直接调用大模型生成教学内容；二是单一智能体生成方案，即仅使用领域知识生成 Agent，不经过学情诊断和审核裁判环节；三是传统标准化课件方案，即使用固定的课程讲义和习题，不进行个性化调整。三种基线方案分别代表了当前主流的 AI 教学工具、简易智能体系统和传统在线学习平台的典型实现，均使用相同的测试问题集和评估标准。

5.4 实验结果

（1）内容质量实验结果（见表 1）：AI 生成教学内容幻觉率低于 3%，远低于通用大模型直出的幻觉水平（约 12%）；知识点匹配准确率较通用大模型直出方案提升 15%–21%，生成内容与课程教材的契合度显著提高。六维学情评估模型的能力判定与人工评估结果的一致性达 89%，可有效识别学生知识盲区并推送匹配的补强资源。混合 RAG 检索模式下的 Top3 相关文档命中率达 98%以上，确保 AI 生成内容有权威教材来源支撑。审核裁判 Agent 的多维度校验机制有效拦截

了不符合要求的生成内容，内容首次通过率约为 85%，未通过内容自动回流生成节点重新优化后，二次通过率可达 95%以上。消融实验表明，移除混合检索后幻觉率上升至 8.7%，移除审核裁判后幻觉率上升至 6.3%，验证了各模块对内容质量的贡献。

表 1.不同方案内容质量与性能对比

对比方案	幻觉率	知识点匹配率	检索延迟（ms）	生成效率提升
通用大模型直出	12.1%	72.3%	-	-
单一智能体生成	6.3%	81.5%	165	35%
传统标准化课件	-	-	-	-
LearnWeave（本文）	2.8%	90.6%	172	62%

（2）分层个性化实验结果：基础薄弱型画像获得的学习资源侧重基础概念讲解、分步代码注释和入门级习题，讲义中基础知识点占比约 70%；正常进度型画像获得标准难度的理论和实操内容，基础与进阶内容比例约为 5：5；高水平型画像获得综合案例、进阶代码和拓展习题，进阶内容占比约 65%。三类画像的学习资源在内容深度、代码注释详细程度、视频演示时长和习题难度梯度上均呈现显著差异。学生学习效率对比实验表明，使用个性化学习路径的学生在相同学习时间内的知识点掌握率比使用标准化课件的学生高出约 23%。

5.5 功能测试

系统测试覆盖后端接口、前端交互、知识检索、核心性能和角色权限边界。学生端学习画像、学习空间、实验中心、题库练习、AI 答疑、学习评估等全链路功能运行正常，三类不同画像的学习者可获得差异化难度的学习资源。学生端形成“画像诊断-资源学习—实操练习—答疑补强—效果评估”完整学习闭环，学习空间支持讲义、思维导图、代码案例、教学视频、分层练习五类资源的无缝切换。教师端班级学情监控、成绩管理、Agent 任务调度功能稳定，可批量查看班级共性薄弱知识点。管理员端全域账号管理、知识库运维、内容质量评测、Agent 运行监控功能均符合设计预期。三角色权限隔离严格，未授权访问均被拦截。

5.6 性能测试

课程材料检索延迟稳定低于 180ms，混合检索模式下的 Top3 相关文档命中率达 98%以上。多模态素材全自动生成效率较人工制作提升 60%以上，单知识点五类资源的生成时长控制在 3 分钟以内，其中文本、导图、代码可实现秒级生成，音频与视频生成效率提升最为显著。传统人工制作单知识点的五类教学资源平均需要 2-3 小时，系统自动生成将时间压缩至 3 分钟以内。在线代码实验批改响应时间低于 2s，支持 Python、Scala 等多语言代码的语法检查、运行结果校验与错误定位。系统接口并发支持 50 用户同时在线操作，响应时间无明显衰减，满足常规班级教学使用需求。

5.7 对比实验

将系统生成内容与通用 AI 工具、传统标准化课件进行对比（结果见表 2），结果表明：LearnWeave 生成的教学资源在知识点匹配度、难度适配性、内容专业性上均优于通用 AI 工具；在个性化程度、多模态丰富度上显著优于传统标准化课件。与通用大模型直出方案相比，本系统的知识点匹配准确率提升 15%–21%，幻觉率从约 12%降低至 3%以下，内容专业性评分（满分 10 分）从 6.2 分提升至 8.7 分；与单一智能体生成方案相比，本系统的内容专业性和难度匹配度提升约 18%，审核通过率提升约 25%，多模态内容一致性评分从 7.1 分提升至 9.0 分；与传统标准化课件相比，本系统的个性化程度和学生学习效率提升约 23%，多模态资源丰富度显著优于纯文本课件。

表 2.系统与同类产品多维度对比

对比维度	通用 AI 工具	单一智能体系统	传统标准化课件	LearnWeave（本文）
知识点匹配率	72.3%	81.5%	-	90.6%
幻觉率	12.1%	6.3%	-	2.8%
个性化程度	低	中	无	高
多模态丰富度	中	中	低	高
角色权限管理	无	简单	有	完善

对比传统标准化线上教学平台，本系统具备三智能体协同个性化多模态资源自动生成、六维量化学情评估、本地 RAG 知识库防幻觉、在线代码实验自动批改四大独有能力，不存在内容同质化、AI 生成不可控、实操训练缺失痛点。对比通用 AI 答疑工具，本系统绑定校内课程知识库，区分学生/教师/管理员完整教学管理体系，拥有完整学习闭

环，并非单一问答工具，适配高校完整教学流程。对比单一自适应刷题系统，本系统覆盖讲义、视频、实操、答疑、学情全链路，不局限于习题训练，兼顾理论学习与代码实操。用户调研显示，参与测试的 30 名学生中，27 名表示愿意继续使用本系统进行学习，23 名认为系统推送的资源比传统课件更适合自己的学习节奏；参与测试的 5 名教师均认为系统能够有效减轻备课和学情分析负担，生成的教学内容质量达到课堂使用标准。

6.结论与未来工作

本文设计并实现了 LearnWeave 多智能体个性化知识生成系统，通过 LangGraph 三智能体协同调度、混合 RAG 检索增强、一源多生多模态自动生成、六维学情量化评估核心技术，形成覆盖学习者诊断、资源生成审核、代码实践、智能问答和学习效果评价的完整教学闭环。系统可降低教师重复备课与学情分析负担，有效抑制大模型幻觉，并为不同基础学生提供差异化学习材料，适配高校计算机类课程分层个性化教学需求。系统严格区分学生自主学习、教师教学管理、管理员平台运维三类角色，构建从画像诊断、多模态资源学习、代码实验实操、智能答疑、效果评估到教学管控、全域运维的完整智慧教育业务闭环。

实验结果表明，三智能体串行协作机制可实现教学资源生成的全流程自动化与可追溯化，学情诊断 Agent、领域知识生成 Agent、审核裁判 Agent 的分工协作确保了生成内容的专业性和可靠性，任务全流程可视化可追溯；混合 RAG 防幻觉机制显著提升了生成内容的准确性与教材契合度，AI 生成教学内容幻觉率低于 3%，知识点匹配准确率较通用大模型直出方案提升 15%–21%，Top3 相关文档命中率达 98%以上；多模态自动生成技术大幅降低了教学资源制作成本，单知识点五类资源的生成时长控制在 3 分钟以内，生成效率提升超过 60%，教师备课工作量显著降低；六维评估模型实现了学情的综合量化刻画，能力判定与人工评估结果的一致性达 89%，为分层教学提供了数据支撑，评估结果与学习资源动态联动实现了评估即补强。技术层面，本系统将多智能体协同、检索增强生成、多模态教学素材生成技术落地高校教育场景，提供轻量化私有化部署方案，无需依赖第三方教育 SaaS 平台，降低院校数字化教学落地成本，为 AI 自适应学习系统提供“诊断—生成—审核—评估”完整技术实现范例。社会层面，系统实现因材施教数字化落地，针对不同基础学生推送适配学习内容，提升自主学习效率；自动化备课、学情分析减轻教师重复性工作；本地知识库机制规避通用大模型知识错误问题，保障高校教学内容专业准确，助力智慧校园数字化建设。

由于当前系统的课程知识库主要覆盖计算机类专业，学科适配范围存在一定局限性，大语言模型的推理能力和多模态理解能力仍有提升空间，因此未来的工作将从以下方面展开：1）功能迭代：拓展小组协作学习、教师直播授课、实验报告自动评分模块，增加移动端适配前端页面，支持学生在手机和平板设备上进行碎片化学习；2）技术升级：引入多模态大模型，支持图片、代码截图答疑，实现图文混合输入的智能问答；优化向量检索速度，支持百万级课程文档检索，引入更高效的近似最近邻搜索算法；探索知识图谱与 RAG 的融合，提升复杂知识点的关联推理能力；3）场景拓展：适配中小学理科、经管类专业课程，扩充学科知识库模板，建立不同学科的知识点体系和教学资源生成规范；4）运维优化：增加系统自动化备份、一键部署脚本，简化院校本地部署操作，提供 Docker 容器化部署方案和 Kubernetes 集群部署方案；5）安全升级：增加操作日志审计、教学数据脱敏导出功能，进一步提升隐私保护能力，引入联邦学习技术，在保护学生隐私的前提下实现跨校学情模型协同优化。

参考文献

[1] Wang X, Xu S, Zhang Y, et al. Exploring the impact of artificial intelligence application in personalized learning environments: thematic analysis of undergraduates’ perceptions in China[J]. Humanities and Social Sciences Communications, 2024, 11:1644. DOI:10.1057/s41599-024-04168-x.

[2] Bayly-Castaneda K B, Ramirez-Montoya M S, Morita-Alexander A. Crafting personalized learning paths with AI for lifelong learning: a systematic literature review[J]. Frontiers in Education, 2024, 9:1424386.

DOI:10.3389/feduc.2024.1424386.
[3] Och F J, Ney H. Statistical multi-source translation[C]//Proceedings of MT Summit. 2001, 8:253-258.
[4] Zoph B, Knight K. Multi-source neural translation[C]//Proceedings of NAACL-HLT. 2016:30-34. DOI:10.18653/v1/N16-1004.
[5] Garmash E, Monz C. Ensemble learning for multi-source neural machine translation[C]//Proceedings of COLING 2016. 2016:1409-1418.
[6] Dabre R, Cromieres F, Kurohashi S. Enabling multi-source neural machine translation by concatenating source sentences in multiple languages[C]//Proceedings of Machine Translation Summit XVI. 2017.
[7] Libovický J, Helcl J, Mareček D. Input combination strategies for multi-source Transformer decoder[C]//Proceedings of the Third Conference on Machine Translation: Research Papers. 2018:253-260. DOI:10.18653/v1/W18-6326.
[8] Li H, Chong Y, Stepputtis S, et al. Theory of mind for multi-agent collaboration via large language models[C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023:180-192. DOI:10.18653/v1/2023.emnlp-main.13.
[9] Chen J, Saha S, Bansal M. ReConcile: Round-table conference improves reasoning via consensus among diverse LLMs[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. 2024:7066-7085. DOI:10.18653/v1/2024.acl-long.381.
[10]Ha T L, Niehues J, Waibel A. Toward multilingual neural machine translation with universal encoder and decoder[C]//Proceedings of the 13th International Conference on Spoken Language Translation. 2016.
[11]Johnson M, Schuster M, Le Q V, et al. Google's multilingual neural machine translation system[J]. Transactions of the Association for Computational Linguistics, 2017, 5:339-351. DOI:10.1162/tacl_a_00065.
[12]Blackwood G, Ballesteros M, Ward T. Multilingual neural machine translation with task-specific attention[C]//Proceedings of the 27th International Conference on Computational Linguistics. 2018:3112-3122.
[13]Sachan D, Neubig G. Parameter sharing methods for multilingual self-attentional translation models[C]//Proceedings of the Third Conference on Machine Translation: Research Papers. 2018:261-271. DOI:10.18653/v1/W18-6327.
[14]Platanios E A, Sachan M, Neubig G, Mitchell T. Contextual parameter generation for universal neural machine translation[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:425-435. DOI:10.18653/v1/D18-1039.
[15]Huang W, Lapata M, Vougiouklis P, et al. Retrieval augmented generation with rich answer encoding[C]//Proceedings of IJCNLP-AACL 2023. 2023:1012-1025. DOI:10.18653/v1/2023.ijcnlp-main.65.
[16]Xu S, Pang L, Yu M, et al. Unsupervised information refinement training of large language models for retrieval-augmented generation[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. 2024:133-145. DOI:10.18653/v1/2024.acl-long.9.
[17]Es S, James J, Espinosa Anke L, Schockaert S. RAGAs: Automated evaluation of retrieval augmented generation[C]//Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations. 2024:150-158. DOI:10.18653/v1/2024.eacl-demo.16.