

图文协同：基于多智能体协作的物理多模态习题生成研究

Image-Text Synergy: A Multi-Agent Collaborative Approach for Multimodal Physics

Question Generation

翁基栋^{1*}, 陈韬宇¹, 蒋杭奇¹, 李浩君¹¹浙江工业大学教育学院

* wjd_edu@163.com

【摘要】针对通用大模型在物理习题生成中存在的“图文语义不一致”问题，本研究提出图文协同生成框架 Physics-MAC。该框架构建了由命题、绘图与审核智能体组成的协同工作流，创新性提出“代码绘图+大模型文生图”混合绘图策略，通过智能路由实现对强几何约束与现实情境类题目的差异化精准绘图。实验表明，该系统在图文一致性（4.18/5.00）与科学严谨性（4.35/5.00）上显著优于基线模型，资源可用率达 74.0%。本研究验证了多智能体协作在消解跨模态逻辑冲突上的有效性，为理工科精准教学资源的自动化生产提供了新范式。

【关键词】多智能体系统；习题生成；图文协同；智能教学系统；精准教学

Abstract: To address the challenge of "image-text semantic misalignment" prevalent in physics problem generation by general Large Language Models (LLMs), this study proposes Physics-MAC, an image-text synergistic generation framework. This framework constructs a collaborative workflow comprising Question Design, Illustration, and Quality Review Agents. It innovatively introduces a hybrid rendering strategy combining "code-driven plotting" with "generative text-to-image" synthesis. By utilizing an intelligent routing mechanism, the system achieves differentiated and precise rendering for problems requiring either strong geometric constraints or realistic contextual scenarios. Experimental results demonstrate that the proposed system significantly outperforms baseline models in terms of Image-Text Consistency (4.18/5.00) and Scientific Rigor (4.35/5.00), achieving a Resource Usability Rate of 74.0%. This study validates the effectiveness of multi-agent collaboration in resolving cross-modal logical conflicts, offering a novel paradigm for the automated production of precision teaching resources in STEM disciplines..

Keywords: Multi-Agent Systems; Problem Generation; Text-Image Synergy; Intelligent Tutoring Systems; Precision Teaching

1. 前言

在人工智能赋能教育的时代背景下，教育范式正加快从标准化的大规模教学向以学习者为中心的精准教学转型。精准教学强调对学生的认知状态进行诊断，并即时提供适配的教学资源与干预策略（Kong et al., 2024）。在此过程中，高质量、多层次且针对性强的习题资源是实现个性化教学的关键载体。特别是在物理学科教学中，为了帮助学生克服概念迷思或巩固薄弱知识点，教师往往需要大量的变式题——即考查逻辑相同但情境参数不同的题目。

然而，物理习题具有显著的多模态特征，其表征形式依赖于文本与图像的双重编码，严谨的题干描述必须与精确的物理情境图（如受力分析图、运动轨迹图）相互配合。这种特性使得高质量物理题库的建设面临严重的瓶颈：人工命题耗时耗力，难以满足大规模个性化教学中对海量变式资源的实时需求。尽管生成式人工智能的兴起为教育资源的自动化生产提供了新契机（卢宇等，2023），但如何生成符合学科逻辑且图文高度一致的物理习题，仍是当前智能教育应用的一大挑战。

近年来，以 ChatGPT 为代表的大语言模型（Large Language Models, LLMs）在自然语言处理任务中展现了卓越的性能，并开始被广泛应用于自动问题生成（Automatic Question Generation, AQG）领域（Wang et al., 2024）。然而，现有的通用大模型在处理物理学科任务时，往往暴露出明显的局限性。

首先，物理学科要求极高的逻辑严密性。由于缺乏特定领域的学科教学知识和物理出题、绘图等技巧，通用大模型常产生幻觉现象。例如，生成的题干描述“物体在粗糙斜面上滑下”，而生成的配图却未标注摩擦系数或受力方向错误，导致严重的“图文语义不一致”（Mulla & Gharpure, 2023）。其次，现有的多模态生成模型多基于像素扩散生成，直接利用模型生成难以精确控制图像中的几何参数（如特定的角度数值或矢量长度），导致生成的图像缺乏科学性，无法直接应用于严谨的教学场景。

针对上述挑战，单一的大模型难以同时兼顾教育目标设计、物理逻辑推理与精确图像绘制

等多重任务。受多智能体系统 (Multi-Agent Systems, MAS) 在解决复杂协同任务中展现出的优势启发 (Guo et al., 2024; Li et al., 2023), 本研究提出了一种“图文协同”的物理习题生成框架。

该框架模拟人类教研组的专业分工模式, 构建了由“命题设计智能体”、“配图绘制智能体”与“质量审核智能体”组成的协作系统。本研究的主要贡献包括:

提出了一种面向物理学科的图文协同习题生成方案。针对通用大模型在物理空间逻辑上的“幻觉”问题, 本研究通过多智能体角色扮演将复杂的跨模态生成任务解耦, 建立了从文本逻辑到视觉表征的“思维链”协同工作流, 显著提升了生成内容的逻辑一致性。

创新性地构建“代码绘图与大模型生成”的混合视觉表征策略。针对物理习题的差异化图文特征, 本研究设计了双轨制绘图机制, 有效解决了单一技术路线在“科学严谨性”与“情境丰富度”之间的权衡难题。

融合布鲁姆教育目标分类学实现习题难度的可控生成。在命题智能体的提示工程中嵌入教育心理学理论, 实现了从“记忆/理解”到“分析/评价”的认知层级跃迁, 为精准教学提供了分层化、个性化的题库资源支持。

2. 研究现状

自动问题生成 (AQG) 技术经历了从规则模板到深度神经网络, 再到预训练大模型的演进历程 (Mulla & Gharpure, 2023)。早期基于模板的方法虽然保障了语法正确性, 但形式僵化; 深度学习模型提升了文本流畅度, 却难以把控逻辑考点。近年来, 大语言模型 (LLM) 的介入显著拓展了 AQG 的应用边界。LLM 在多语言生成及开放式问答中展现了卓越性能, 使大规模习题资源的自动化建设成为可能 (Wang et al., 2024)。不少基于提示词工程的物理、数学等学科的文本习题生成方案、系统应运而生 (Luo et al., 2024)。

然而, 现有 AQG 研究主要集中于通用领域的纯文本生成。卢宇等人 (2024) 在实证研究中发现, 通用模型在处理物理、数学等强逻辑学科时, 常因缺乏学科教学知识而产生事实性错误。更关键的是, 针对理工科习题“图文协同”难点的研究有待探索。尽管图像生成技术发展迅速, 但现有文献表明, 单一模型难以实现对物理几何参数 (如角度、矢量比例) 的精确控制。目前鲜有研究能够成功解决 AQG 任务中“文本逻辑与科学绘图精准对齐”的技术难题, 这一缺失直接制约了智能题库在精准教学中的落地。

针对单一模型在处理跨模态、长链条推理任务时的能力局限, 多智能体系统 (MAS) 提供了新的解决视角。通过角色扮演与任务分解, MAS 能够模拟人类群体的协作智慧, 显著提升复杂逻辑任务的解决成功率 (Guo et al., 2024)。通过 CAMEL 框架证实, 智能体之间的自主交互与辩论机制可以有效激发深层推理, 修正单一视角的逻辑谬误 (Li et al., 2023)。

在教育应用层面, EduAgentQG 已初步探索了利用多智能体工作流生成个性化题目, 验证了“分工协作”模式在提升题目教育针对性上的有效性。然而, 该研究主要聚焦于文本维度的优化, 尚未涉足复杂的跨模态 (图文) 协同 (Jia et al., 2025)。现有文献表明, 多智能体协作机制有望成为解决物理学科“图文对齐”难题的关键路径。本研究将重点构建面向物理学科的图文协同生成机制, 在多智能体分工协作的作用下, 实现文本逻辑与习题配图的精准对齐与自动化生成, 从而达到破解“图文语义不一致”难题, 满足精准教学对高质量变式资源的个性化需求, 为物理教师命题绘图工作增效减负的目的。

3. 基于多智能体协作的物理多模态习题生成

3.1 多智能体综合方案

本研究构建了名为 Physics-MAC (Physics Multi-Agent Collaboration) 的图文协同物理习题生成系统。该系统架构模拟了人类物理教研组的专业分工模式, 通过大语言模型的角色扮演机制, 将复杂的跨模态生成任务解耦为由命题设计智能体、配图绘制智能体与质量审核智能体共同维护的协同工作流。配图绘制智能体使用 doubao-seedream-3-0-t2i 作为底座, 其余智能体均用 doubao-seed-1-6 作为底座。

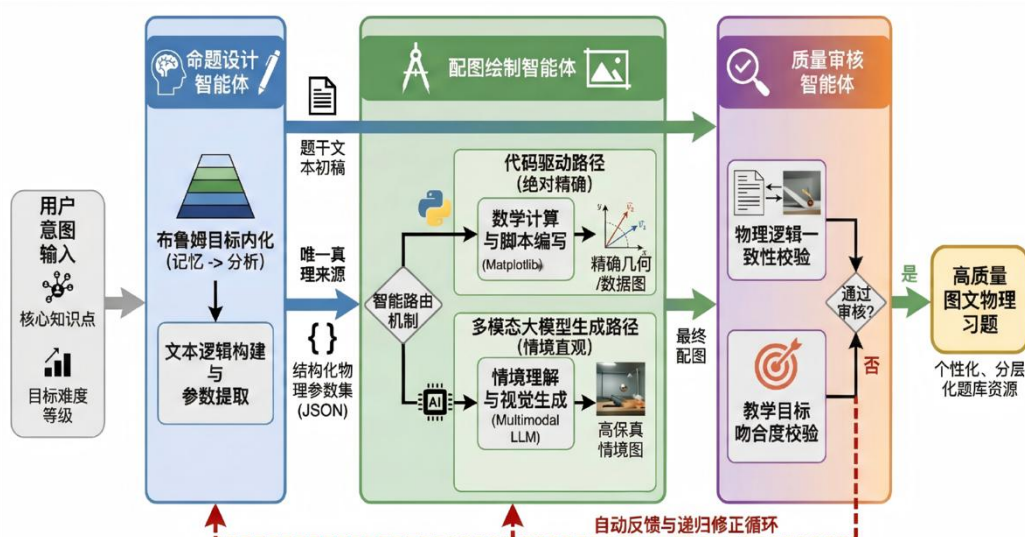


图1 Physics-MAC 多智能体系统框架

在系统的上游，命题设计智能体作为“教育目标的转化者”，内置了布鲁姆教育目标分类学的认知提示模版，负责根据用户意图确立习题的考查难度与知识点，并从非结构化的文本中提取出精确的物理参数集合，以结构化数据（JSON）的形式向下游传递，确保了后续生成任务具有唯一的“真理来源”。

针对物理学科习题表征的多样性，本研究在配图绘制智能体环节创新性地设计了“代码+生成式”混合绘图策略，并通过智能路由机制实现了绘图路径的动态优选。该路由机制的核心判别边界在于题目题干中是否包含决定物理状态的显性几何约束参数（如具体角度、坐标比例、严格的力学矢量关系）。基于此边界，智能体执行精准的路径分发，若包含上述约束，则硬性路由至“代码驱动”路径，利用大模型将物理参数转化为 Python（Matplotlib）绘图脚本，凭借数学计算的确定性彻底消除几何参数误差，确保配图的科学保真；若题目仅需背景情境渲染（如生活化物理场景），则路由至“多模态大模型生成”路径，直接生成高保真度的情境图像，以增强题目的直观性与趣味性。值得一提的是，对于同一习题中并存多种类型配图需求（如需同时生成受力原理图与真实情境图）的复杂情况，该路由机制能够将多图任务自动解耦，对各个图像子任务进行独立判别与并行分发。这种双轨制策略在保证物理科学严谨性的同时，有效兼顾了教学情境的丰富度以及复杂习题的多图表征需求。

为了全面保障最终输出资源的科学严谨性与教学适配度，系统在末端引入了质量审核智能体作为图文一致性与教学目标的双重把关者。该智能体采用多维校验机制，重点执行两层核心审核。首先是物理逻辑一致性校验，核查视觉图像与题干文本在物理情境上是否存在冲突。同时，还要进行教学目标吻合度校验，回溯比对生成的习题是否精准覆盖了用户意图层设定的核心知识点，以及是否符合预设的布鲁姆认知难度层级。

一旦检测到物理逻辑错误或教学目标偏离，系统将触发自动反馈循环，根据错误类型生成具体的修正指令并回传至相应的前序智能体进行迭代优化，直至各项指标均通过审核。通过这种“串行生成-递归修正”的闭环机制，Physics-MAC 系统实现了从教育目标设计到高质量图文资源产出的全流程自动化与精准化。

3.2 提示词编写方法

提示词质量对于大语言模型在特定垂直领域的生成效果有着显著的影响。作为构建多智能体系统的关键变量，不同类型和构造的提示词对模型的理解深度与输出规范性产生截然不同的效果。本研究采用“角色扮演”与“结构化约束”等提示工程策略，通过赋予每个智能体明确的专家身份与任务边界，确保其行为符合物理学科的严谨逻辑。

针对不同的任务复杂度，本研究采取了差异化的提示构建方法。对于命题设计与质量审核，主要采用零样本学习以激发模型的通用推理能力；而对于任务边界较为模糊的绘图路径

判别任务，则引入少样本学习策略。在复杂分类任务中提供具体的范例能显著提升大模型对规则的理解力（Jia et al.,2025）。因此，本研究在执行智能路由机制的智能体提示词中嵌入了具体的判别范例，以实现“代码绘图”与“生成式绘图”路径的精准分发。各核心智能体的详细提示词设计见表 1。

4.实验结果

为了验证 Physics-MAC 系统在物理习题生成任务中的实际效能，本研究构建了包含 50 个高中物理选择填空题核心考点的测试数据集，并设定了差异化的认知难度目标进行对比实验。其中，基线组（Baseline）采用主流的通用大模型直接生成方案，即利用 doubao-seed-1-6 通过结合思维链，少样本学习等提示词撰写技巧设计提示词生成题干及图像描述，并直接调用 doubao-seedream-3-0-t2i 生成配图，以此代表当前单体大模型在未经过多智能体增强下的平均水平。

首先需要验证图文协同机制在精细化生成中的判别效能，本研究对智能路由 Agent 的路径分发准确率进行了独立评估。不同于传统的整题分类模式，本系统采用图像任务级的路由策略，即首先根据题目考点自动解析出所需的图像数量及内容描述，再针对每一张独立的图像描述进行路径规划。以测试集中的 50 道核心考点题为基准，系统将其自动解耦为 64 个独立的绘图任务。本研究邀请了三三位经验丰富的物理教师及教育研究者组成专家评审团，对各图像描述的理想表征形式进行了人工标注。其中 40 个任务涉及严格几何约束（需激活 Path A），24 个任务侧重现实情境渲染（需激活 Path B）。实验统计显示，智能路由 Agent 成功对其中 61 个绘图任务执行了正确的分发指令，任务级分类准确率达 95.3%。这一结果证实，该机制具备优异的多图拆解与精准适配能力，即使面对单一题目需要同时呈现原理示意图与真实情景图的复杂需求，也能通过并行路由机制实现图文资源的精准供给。

进一步，为了确保评估结果的客观性与科学性，专家评审团从图文一致性、科学严谨性及教学目标契合度三个维度对所生成的选择填空习题进行量化评分（满分 5 分）。在正式评估前，本研究开展了评分标准校准的预实验。三位专家首先对从测试集中随机抽取的 10 组样本进行独立试评，随后通过组内讨论明确了各维度的具体给分细则，以消解主观理解偏差。

表 1 Physics-MAC 多智能体协作系统的核心提示词设计

智能体角色	提示词设计架构
命题设计智能体策略：零样本+结构化输出	请您扮演一名资深的高中物理命题专家。您精通布鲁姆教育目标分类学与物理课程标准。 1.任务目标：根据用户输入的【知识点】与【难度等级】，生成一道逻辑严密、表述清晰的物理习题。 2.核心约束：必须从生成的题干中提取出所有决定物理情境的关键参数（如质量、角度、摩擦因数、运动状态等），并将其转化为标准的 JSON 格式，作为后续绘图的唯一“真理来源”。 3. 输出格式：{"Question_Text": "...", "Bloom_Level": "...", "Physics_Parameters": {"model_type": "slope", "mass": "2kg", "angle": "30_deg", "friction": true, "vectors": ["gravity", "support_force"]}}
智能路由机制（智能体）策略：少样本学习	请您扮演一名多模态任务调度专家。 您的任务是根据输入的【物理参数】与【题目描述】，判断该题目最适合的绘图方式。 分类标准： Path_A (代码驱动): 适用于几何约束强、需定量分析、涉及数据图表（如受力分析图、电路图、v-t 图）的题目。Path_B (生成式驱动): 适用于侧重生活情境展示、几何精度要求低、需强调视觉直观性（如宇航员出舱、风力发电机外观）的题目。 【范例 1】输入：{"题目": "一物块静止在倾角为 30°的粗糙斜面上..." }输出：{"Route": "Path_A", "Reason": "涉及精确角度与受力矢量，需代码精确绘图"} 【范例 2】输入：{"题目": "如图所示，我国神舟十九号返回舱在夕阳下成功着陆..." }输出：{"Route": "Path_B", "Reason": "侧重现实情境渲染，无严格几何定量约束"} 【范例 3】输入：{"题目": "某质点运动的速度-时间图像如图所示..." }输出：{"Route": "Path_A", "Reason": "属于标准科学图表，需 Matplotlib 绘制"} 当前任务： 请对以下输入进行判别：【输入数据】
配图绘制智能体策略：条件分支执行	请您扮演一名精通 Python 可视化的绘图工程师与视觉艺术家。 根据路由智能体的分发指令执行相应操作： 1. 若指令为 Path_A：读取 JSON 参数，编写 Python (Matplotlib) 代码。要求：代码中的几何参数必须与 JSON 数据绝对一致，并自动添加物理标注。 2.若指令为 Path_B：将物理情境转化为高质量的图像生成提示词。要求：包含风格描述、主体细节及光影构图，调用模型生成高保真图像。 输出：可执行代码块 或 图像生成提示词。
质量审核智能体策略：双重校验与反馈	请您扮演一名严谨的物理教研组长。 负责对生成的“题干-配图”进行终审。 1. 物理逻辑校验：对比题干与配图（或代码逻辑），检查是否存在冲突。 2. 教学目标校验：评估题目是否达到了预设的布鲁姆认知难度，知识点覆盖是否准确。 反馈机制：若通过，输出 PASS。若不通过，输出 FAIL 并给出具体修改建议：【修改建议】。

为验证专家评审团评分的一致性，本研究采用了肯德尔和谐系数 (Kendall's W) 对专家评分进行信度检验。统计结果显示，三位专家在图文一致性、科学严谨性及教学目标契合度三个维度上的评分一致性系数分别为 0.82、0.86 和 0.81 ($p<0.01$)，均表明评分者之间具有极高的一致性。基于此，本研究最终采用三位专家评分的算术平均值作为每道习题的最终得分，以最大限度地降低个体主观偏差对实验结果的影响。

如表 2 数据所示, Physics-MAC 系统在各项关键指标上均优于基线模型。在核心的“图文一致性”与“科学严谨性”维度, 本系统的评分分别达到了 4.18 与 4.35, 相比基线提升明显。这一性能优势主要验证了“代码+生成式”混合视觉表征策略的有效性, 该策略通过代码的精确计算, 在很大程度上减少了基线模型在处理受力分析、运动轨迹等强几何约束题目时常见的“几何幻觉”现象(如角度偏差或矢量缺失)。相比之下, 基线组虽然在文本生成的教学目标契合度上表现尚可(3.85 分), 但受限于单一模型的跨模态对齐能力, 其生成的图文资源常存在逻辑冲突, 导致最终的资源可用率仅为 42.0%。

此外, Physics-MAC 系统达到了 74.0% 的资源可用率, 证实了多智能体协作机制在实际应用中的可行性。实验分析表明, 虽然仍有部分复杂情境题目需要人工微调, 但大部分生成的习题能够通过质量审核智能体的“双重校验+递归修正”机制达到直接可用的标准。综上所述, 实证数据表明 Physics-MAC 系统能够较为稳定地输出符合物理科学规范且匹配教学目标的高质量习题资源, 为个性化题库建设提供了有效的技术支撑, 图 2 为系统生成图文结合的物理习题的一个实例。

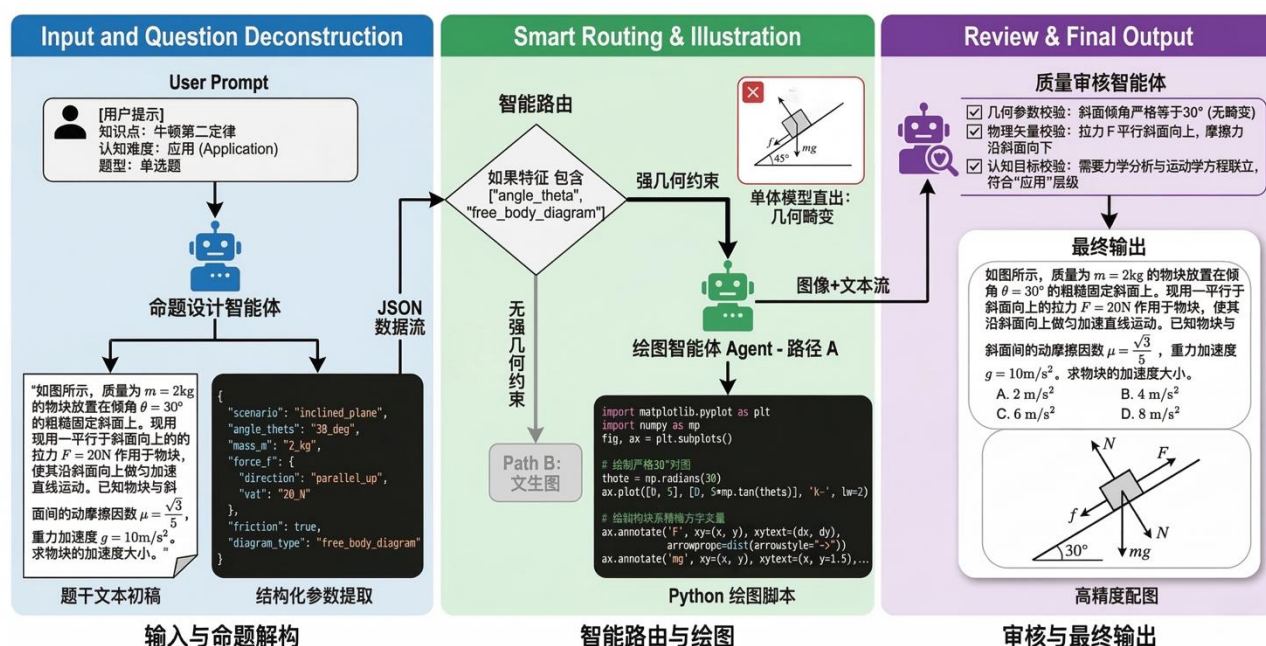


图 2 习题生成系统 workflow 实例

为了客观审视系统在教学应用中的边界, 本研究对评估中未达到直接可用标准的 13 个失效样本进行了深度剖析与量化归类。统计显示, 当前系统的生成局限主要集中在以下三类:

(1) 物理空间拓扑构建错误 (5 例, 占比最高): 尽管系统在常规物理情境下表现稳定, 但在处理涉及多体耦合运动的高阶认知题目时, 大语言模型在空间逻辑推理上的局限性依然凸显。具体案例显示, 在处理“多物块叠放的受力分析”时, 系统虽能在文本中正确罗列受力情况, 但生成的配图却未能准确识别物体间的空间接触面边界, 常将摩擦力或弹力矢量绘制在脱离物体的悬空区域, 或是方向与相对滑动趋势完全反转的视觉谬误。

(2) 代码编译与调用异常 (4 例): 在利用代码驱动路径生成极端复杂的几何图形时, 偶发底层 API 参数不兼容或坐标系越界问题, 导致 Python 脚本编译失败而无法输出图像。

(3) 生成式绘图细节失控 (4 例): 在利用大模型进行现实情境渲染时, 尽管整体氛围逼真, 但偶发局部主体部位的线条粘连或多余元素生成的视觉冗余。

从教育心理学角度来看, 上述深层的图文语义冲突若未被及时修正, 极易在教学过程中引发学习者的认知失调, 甚至诱发顽固的科学迷思。这一发现深刻表明, 在应对极高严谨性的理工科教学资源生产时, 科学性与直观性仍面临难以完美兼顾的挑战。当前基于生成式人工智能的自动化出题框架, 仍亟需引入外部几何约束引擎作为技术支架, 以切实保障学科逻辑

辑的准确性与教育效度。

表 2 不同生成系统的性能指标对比 (N=50)

模型架构	图文一致性评分	科学严谨性评分	教学目标契合度	资源可用率
Baseline	3.15 ± 0.42	3.42 ± 0.38	3.85 ± 0.25	42.0%
Physics-MAC	4.18 ± 0.21	4.35 ± 0.18	4.52 ± 0.24	74.0%

注：“资源可用率”是指在专家评审中，习题的所有维度评分均达到及格标准 (≥ 3 分)，且无需大幅修改即可直接投入教学使用的样本比例。

5. 结语

本研究提出的 Physics-MAC 框架不仅突破了单一模型在理科多模态生成中的能力瓶颈，也为精准教学提供了可落地的技术方案。其核心价值在于：一方面，通过布鲁姆认知层级的可控生成，能够动态产出符合特定教学目标的变式题，为个性化学习提供了分层化的题库资源支持；另一方面，所提出的“代码绘图”与“大模型文生图”相结合的混合策略，为解决理工科习题中“科学精确性”与“情境直观性”的兼容难题提供了新范式，该思路在逻辑上可进一步推广至化学分子结构生成、地理地貌图绘制等类似学科场景中。

本研究也存在一定的局限性。首先，在样本规模与题型覆盖上，当前测试集仅包含 50 个高中物理核心考点，且主要侧重于选择填空题，尚未全面覆盖实验设计、多过程计算等复杂题型，可能无法完全反映系统在极端长链条推理任务中的表现。另外，目前的 Path A（代码绘图）和 Path B（大模型文生图）虽然能覆盖绝大多数基础教学中的静态配图需求，但在处理极度复杂的三维不规则物理模型、或需要高度抽象的拓扑结构时，这两种路径依然存在无法完美表征的局限性。

未来研究方向主要包括：1) 构建多分支、多题型的大规模测试集 (Benchmark)，以验证并增强框架的泛化能力；2) 开发基于苏格拉底式引导的交互式解题辅导智能体；3) 引入视频生成模型，突破静态配图局限，实现对平抛、波动等动态物理过程的进阶表征。

参考文献

- 卢宇, 余京蕾, 陈鹏鹤, 李清堂. (2023). 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例. 中国远程教育, 43(4), 24-31.
- 卢宇, 余京蕾, 陈鹏鹤. (2024). 基于大模型的教学智能体构建与应用研究. 中国电化教育, (7), 99-108.
- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., ... & Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. arXiv preprint arXiv:2402.01680.
- Jia, R., Zhang, M., Liu, F., & Wang, X. (2025). EduAgentQG: A multi-agent workflow framework for personalized question generation. arXiv preprint arXiv:2511.11635.
- Kong, S.-C., & Yang, Y. (2024). A human-centered learning and teaching framework using generative artificial intelligence for self-regulated learning development through domain knowledge learning in K-12 settings. IEEE Transactions on Learning Technologies, 17, 1562-1574.
- Li, G., Hammoud, H., Itani, H., Khizder, D., & Ghanem, B. (2023). CAMEL: Communicative agents for "mind" exploration of large language model society. Advances in Neural Information Processing Systems, 36, 51991-52008.
- Luo, H., Weng, J., Lu, X., & Yu, Z. (2024, October). BPE: Exploring the Prompt Framework of Physics Exercises Generated from Bloom's Taxonomy in LLM. In 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 26-29). IEEE.
- Mulla, N., & Gharpure, P. (2023). Automatic question generation: A review of methodologies, datasets, evaluation metrics, and applications. Progress in Artificial Intelligence, 12(1), 1-32.
- Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., ... & Wen, Q. (2026). Large language models for education: A survey and outlook. IEEE Signal Processing Magazine, 42(6), 51-63.