

CoRIDA: 协作反思式多智能体教学设计生成框架

CoRIDA: A Collaborative Reflective Multi-Agent Framework for Lesson Design Generation

张昕宇, 尚俊杰*

北京大学教育学院

zhxy@stu.pku.edu.cn

【摘要】 课时教学设计需在有限课时内实现学习目标、活动与评价的一致性。本文提出协作反思式教学设计智能体框架 CoRIDA: 以协作组织规划、生成、评审与改稿的多视角分工流程, 以反思将评审意见判据化并转译为可执行改稿机制, 并引入课标与教材知识库的检索增强支持。在 N=30 的小学科学同题对照中, CoRIDA 总分均值 35.37 ± 1.96 , 优于 Qwen 直出 (26.87 ± 2.79) 和 DeepSeek 直出 (25.53 ± 3.15), 优势在评价与学习闭环维度尤为突出。

【关键词】 教学设计; 多智能体; 协作; 反思; 小学科学

Abstract: This paper proposes CoRIDA, a collaborative reflective multi-agent framework for lesson design generation that organizes multi-view collaboration and criterion-based reflective revision. In a same-prompt comparison ($N=30$, primary science), CoRIDA achieves higher overall scores with lower variance, with the largest gains on assessment loop quality.

Keywords: instructional design; multi-agent; collaboration; reflection; primary science

1. 引言

教学设计的核心是实现学习目标、活动与评价的一致性 (Wiggins & McTighe, 2011)。大语言模型虽降低了教案起草成本, 但单轮生成在教学法对齐与评价闭环可操作性方面存在系统性不足 (Hu et al., 2024; Lee & Zhai, 2024)。反思性实践理论强调方案从草稿到可用版本依赖基于证据的再框定与修订 (Schön, 2017)。为此, 本文提出协作反思式教学设计智能体框架 CoRIDA, 聚焦两个机制: 以协作组织多视角分工流程, 以反思将评审意见判据化并转译为可执行改稿。

2. 框架设计与方法

CoRIDA 以逆向设计 (UbD) 的“目标—评价—活动”对齐判据 (Wiggins & McTighe, 2011) 与反思性实践理论的“基于证据检视与修订” (Schön, 2017) 为理论基础。在协作维度, 框架将教案生成组织为四类视角的分工流程: 任务规划——显式化课时、班级、资源与课堂组织限制, 形成共享的“设计任务书”; 方案生成——在统一模板下形成结构齐备且活动链可执行的草案; 多通道评审——从目标对齐、学科准确性、课堂可行性与学生体验等角度检视草案; 综合改稿——对多方意见做取舍与排序, 收束为单一可交付版本。

在反思维度, CoRIDA 将评审设计为面向改稿的判据化诊断: 评审意见须指出缺口属于哪个判据维度、为何影响课堂执行, 并转译为教案中的具体改动要求, 如补齐学习证据、明确成功标准与反馈安排; 综合改稿阶段优先修复会导致“不可用”或“闭环失效”的结构性问题, 以“可交付成品”完成收束。原型系统在规划与评审阶段调用课标与教材知识库的检索增强组件, 输出统一为七部分教案模板, 确保可交付、可对照、可评估。

3. 实验与结果

以 N=30 条小学科学课时教学设计任务为题目集，设三组同题对照：Qwen 单轮直出、DeepSeek 单轮直出、CoRIDA 流程生成，输出归一为同一模板，采用四维量规（目标导向 A、学情适切 B、任务链 C、评价闭环 D，各 1—10 分）与解耦大模型辅助评审（Liu et al., 2023）。三组总分均值±SD 分别为 Qwen 26.87±2.79、DeepSeek 25.53±3.15、CoRIDA 35.37±1.96。整体上，CoRIDA 在四个维度与总分上均表现更高且更稳定（见图 1）。

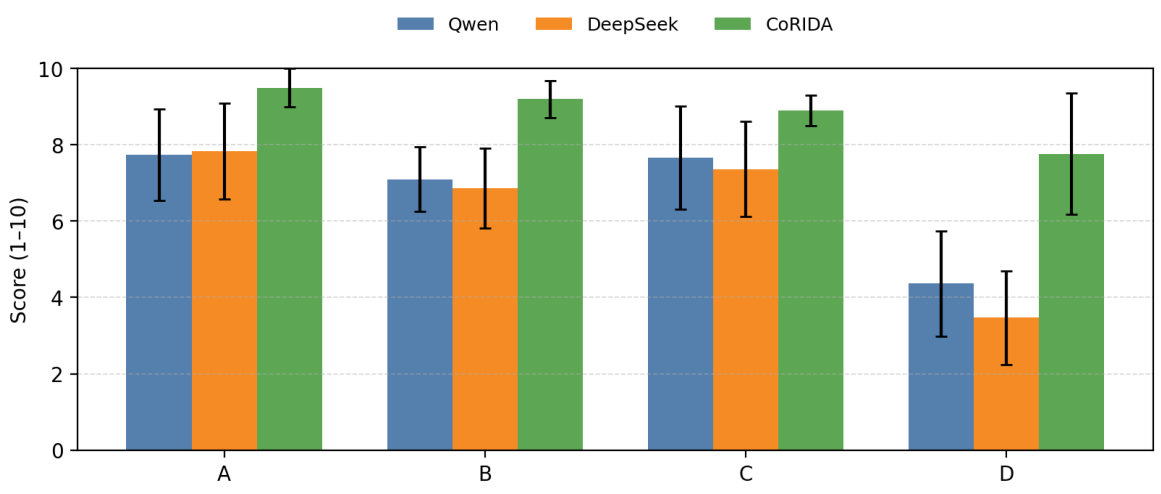


图 1 三组在四个维度上的均值±SD (N=30)

4. 结论

CoRIDA 在小学科学 N=30 同题对照中表现出更高的四维得分与更低波动，说明协同分工与判据化反思有助于提升教学设计质量与输出稳定性。当前验证仍限于小学科学情境，跨学科适用性与真实课堂效果仍待后续研究检验。

参考文献

Hu, B., Zheng, L., Zhu, J., et al. (2024). Teaching plan generation and evaluation with GPT-4. *IEEE Trans. Learn. Technol.*, 17, 1471–1485.

Lee, G. G., & Zhai, X. (2024). Using ChatGPT for science learning: A study on preservice teachers' lesson planning. *IEEE Trans. Learn. Technol.*, 17, 1643–1660.

Liu, Y., Iter, D., Xu, Y., et al. (2023). G-Eval: NLG evaluation using GPT-4 with better human alignment. *arXiv:2303.16634*.

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.

Wiggins, G., & McTighe, J. (2011). *The understanding by design guide to creating high-quality units*. ASCD.