

多智能体人机协同如何支持批判性思维？——基于不同协同配置的实证分析

How Do Multi-Agent Human-AI Collaborations Support Critical Thinking? An Empirical Study Across Collaborative Configurations

楚肖燕^{1*}, 季爽^{1*}, 朱凌芸¹, 钱春雨¹, 翟雪松¹¹浙江大学教育学院

* xiaoyanchu@zju.edu.cn

【摘要】 随着生成式人工智能在教育领域的应用逐渐从工具辅助走向认知协作，单一智能体在促进批判性思维方面仍存在局限。多智能体系统通过多角色分工与反馈，为支持高阶认知提供可能，但不同人机协同配置的影响仍有待检验。本研究以 87 名本科生为研究对象，采用准实验设计考察不同协同配置（单人-多智能体和双人-多智能体协同）对学业表现与批判性思维的影响。结果表明，人机协同模式在知识测验成绩上无显著差异，但双人-多智能体协同的项目产出更优，而单人组在批判性思维意识上表现更佳。滞后序列分析进一步揭示了批判性思维行为的差异特征。

【关键词】 生成式人工智能；多智能体系统；人机协同学习；批判性思维

Abstract: With the increasing adoption of Generative Artificial Intelligence (GenAI) in education, human-AI collaboration has shifted from tool support toward cognitive partnership. Single-agent interactions may suffer from limited perspectives and echo-chamber effects. Multi-Agent Systems (MAS) offer a potential solution, yet the effects of different human-GenAI collaborative configurations remain underexplored. This study examined how different MAS-supported human-GenAI collaboration configurations influence learning performance and critical thinking. A quasi-experimental study involving 87 undergraduates was conducted. Results showed that collaborative learning with MAS led to higher project performance, whereas individual MAS-supported learning enhanced critical thinking awareness. Lag Sequential Analysis further revealed distinct interaction patterns across configurations.

Keywords: Generative artificial intelligence, Multi-agent systems, Human-AI collaborative learning, Critical thinking

1. 引言与研究问题

随着生成式人工智能（Generative Artificial Intelligence, GenAI）的快速发展，人机协同学习逐渐从“工具辅助”阶段迈向“认知协作”阶段。尤其是在高阶认知目标导向的学习情境中，GenAI 进一步被视为能够参与分析、评价与反思的认知伙伴。然而，现有研究多聚焦于单智能体支持的人机互动，其在支持批判性思维方面的有效性仍存在争议。多智能体系统

（Multi-Agent Systems, MAS）提供了新的路径（Hao et al., 2026）。通过引入多个具有不同角色与功能分工的智能体，MAS 能够模拟多视角对话与分布式认知过程，创造更为复杂的认知冲突与意义协商情境。此外，MAS 的效果受到人机协同配置的影响，即以个体还是协作方式与 MAS 互动，以及同伴与 AI 在学习过程中的角色分工与协作关系（Chu & Li, 2025; Tai & Chen, 2024）。现有研究虽然开始关注多智能体支持的人机协同学习，但较少比较不同人机协同配置的差异化影响，且对批判性思维的加工过程关注不足。本研究在 MAS 支持下，系统考察不同协同配置（单人-多智能体和双人-多智能体）对学业表现与批判性思维过程的影响：

RQ1: 在多智能体支持下，不同人机协同配置对学习者的学习表现有何影响？

RQ2: 在多智能体支持下，不同人机协同配置如何影响学习者的批判性思维意识？

RQ3: 在不同协同配置下，批判性思维行为序列模式呈现出哪些特征？

2. 研究背景

2.1. 多智能体系统与批判性思维

批判性思维指的是对“该做什么”和“该相信什么”做出合理判断的能力（Ennis, 1985），涉及问题识别、信息理解、论证分析、证据评价以及解决方案生成等一系列认知过程（Facione, 1990; Murphy, 2004）。通过互动，学习者得以表达自身推理、检视潜在假设并不断修正判断，从而支持批判性思维的发展。GenAI 对学习者的分析性与评价性思维过程起到支架作用，但单智能体交互也面临一些挑战，如“回音室效应”，即确认偏差与视角多样性的不足会削弱学习者进行真实认知投入的机会。

为应对上述局限，多智能体系统被认为是一种架构性解决方案。与单一智能体不同，MAS 通过协调多个具有专门功能的智能体，形成一种分布式探究形态（Hao et al., 2026）。通过模拟多视角对话与复杂问题解决情境，MAS 支持的学习环境能够为学习者创造接触竞争性观点、评估替代性主张并论证自身推理的机会。

2.2. 人机协同配置

除 GenAI 本身的能力，人机协同的效果还受到协同配置方式的显著影响，包括个体或协作式与 AI 进行交互（Chu & Li, 2025; Tai & Chen, 2024）。协同配置规定了学习者、同伴与 AI 的互动组织形式，决定了认知资源分配方式以及信息流动路径。

多智能体系统的引入，进一步拓展并复杂化了人机协同配置。MAS 可以作为共享的认知支架，帮助建立协作团队的整体感并促进对复杂问题的共同理解；此外，也可能作为自主的第三方参与者，通过引入分歧性观点来挑战群体共识，进而激发对既有观点的再评价。目前，系统性比较不同人和多智能体协同配置如何影响学业表现与批判性思维的实证研究仍然稀缺，且对于协作过程中的动态交互机制，亦缺乏深入洞察。

3. 研究方法

3.1. 研究对象与环境

本研究以《现代教育技术》课程作为实验实施情境。该课程具有显著的任务导向与项目驱动特征，为嵌入多智能体支持的人机协同学习活动提供了适切的环境。研究对象为中国一所综合性高校教育学院的本科生，共有 87 名学生参与实验，男性 32 人、女性 55 人。

本研究环境基于 Unity3D 构建前端界面，后端集成 CozeAI 多智能体框架与 DeepSeek 模型，包含技术支持导师、教学设计导师与 AI 伦理导师，承担技术操作指导、教学设计咨询与伦理规范引导等功能。系统依托长时记忆管理与知识库集成等功能，通过角色提示词预设智能体的职责边界与语言风格，形成差异化功能分工与情境化认知支持，按需触发分别提供反馈。基于 RAG 机制接入垂直领域知识库，以保障反馈的专业性与语境相关性。三位虚拟智能体均支持语音、表情与动作等多模态交互。

3.2. 研究流程

本研究采用为期 8 周的准实验设计（如图 1）。实验在微课设计课程中实施，学习者需与同伴或多智能体协同完成一个微课视频的制作。设置三种学习条件：控制组以两人一组完成任务，不使用 AI 支持；单人-多智能体组由学生个体与多智能体系统协；双人-多智能体组则由两名学生在线下协作，共同与多智能体系统交互完成设计活动。三种条件均由同一位教师授课，课程进度、学习目标、教学内容保持一致。干预前后分别实施知识测验与批判性思维意识问卷，结束时提交最终微课设计作品。

图 1

研究实验流程



3.3. 测量工具与数据分析

学业表现由知识测验与期末微课设计项目两部分构成。知识测验包含 10 道选择题，内容涵盖微课设计关键原理与基础软件操作，选自标准化题库。微课设计项目采用包含教学目标（10 分）、教学设计（30 分）、技术实现（20 分）、内容从呈现（20 分）与创意设计（20 分）五个维度的量规进行评价，两位专家独立评分，Cohen's kappa 值为 0.806，一致性良好。

批判性思维意识测量改编自 Hwang 等（2018）的五点李克特量表，Cronbach's α 系数在前后测中分别为 0.774 和 0.924，表明信度良好。

为分析学生在与多智能体系统交互过程中的批判性思维加工模式，基于 Murphy（2004）的框架对对话数据进行编码（见表 1）。两名受训专家独立编码并达到良好一致性（0.848）。

表 1

批判性思维编码框架

阶段	描述	编码
识别	识别或发现已存在的问题、困境或情境	R
理解	探索相关证据、知识、研究、信息与观点	U
分析	澄清问题，组织已知信息，识别未知信息，并将问题分解为基本组成部分	A
评价	对信息、知识或观点进行批判性判断	E
创造	生成新的知识、观点或策略，并加以实施或应用	C

由于数据未完全满足正态分布，使用 Kruskal-Wallis 检验比较学习表现与批判性思维意识方面的差异，采用 Dunn 事后检验并进行 Bonferroni 校正。三组在前测知识成绩和批判性思维意识无显著差异，基线水平相当。采用 GSEQ 软件对编码后的对话进行滞后序列分析。

4. 结果与讨论

在学业表现层面，三组在后测知识测验上未发现显著差异（ $p=0.161$ ），但项目得分组间差异达到显著。事后检验表明，双人-多智能体协同组显著高于对照组（ $p=0.039$ ）和单人-多智能体协同组（ $p=0.044$ ）。这一结果表明，不同学习配置显著影响综合性学习成果，而非知识性测验。双人组的显著优势，显示了同伴协作与多智能体支持的结合在复杂任务情境中具有明显的协同效应。

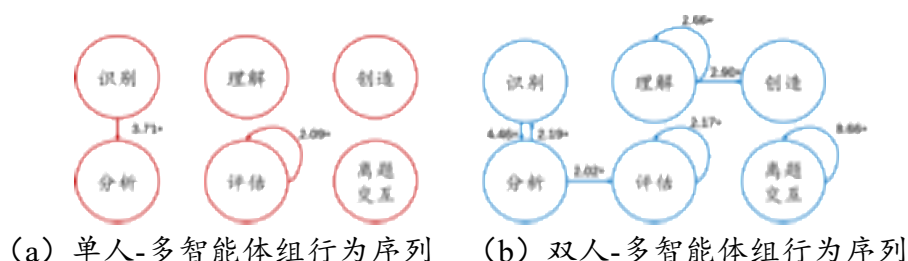
在批判性思维意识层面，三组后测得分存在显著差异（ $p=0.024$ ）。事后比较结果表明，单人-多智能体协同组的得分显著高于双人-多智能体协同组（ $p=0.019$ ）。该结果表明，与协作情境相比，单人-多智能体协同学习为学习者提供了更为集中的认知空间，使其需要独立承担问题理解、方案评估与决策判断的责任。

滞后序列分析结果（图 2）显示，在单人组中，显著的行为转移链主要表现为**识别**→**分析**以及**再评估**，即学习者形成了从问题识别到深入分析的路径，在评估阶段进行检验与反思，反映以个体为中心、强调逻辑推进与评估的批判性思维加工模式。相比之下，双人组的显著行为转移链更加多样化且分散，表现为**识别**与**分析**之间的双向转化、**分析**→**评估**的递进关系，

以及**理解**、**评估**与**离题交互**的循环共现，并最终从**理解**延伸至**创造**。这表明，协作情境中学习者的认知加工并非线性推进，而是通过反复协商、澄清与整合逐步达成共同理解，并生成新的设计方案。结合自我调节学习及共享调节学习理论，双人组的社会协商与任务导向利于整合产出，但可能减少个体元认知反思深度；而单人直面多视角反馈更易激活思维自我监控，但产出整合效率可能较低。

图 2

滞后序列分析结果



5. 研究不足与未来展望

本研究仍存在一定局限。由于样本来源于单一高校课程且采用准实验设计，研究结论的普适性仍需谨慎看待；同时，批判性思维过程分析主要依赖对话数据，尚未整合更多多模态证据。后续研究将从以下几个方面进一步拓展。首先，在研究设计层面引入更具层次性的协同配置与任务复杂度操控，探究不同配置在高、低复杂度任务中的适配性差异。其次，在过程分析层面进一步整合多方法、多模态的数据分析框架，对人机协同路径进行更为精细的刻画。最后，在应用与推广层面可尝试将多智能体人机协同学习模式拓展至不同学科与教育阶段，检验其在不同教学目标与学习群体中的可迁移性。

参考文献

- Chu, X., & Li, Y. (2025). Interacting with a humanized dual-LLM agent: An exploratory study of individual and collaborative modes in language proficiency and critical thinking awareness. *Interactive Learning Environments*, 1–21.
- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational leadership*, 43(2), 44–48.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction : research findings and recommendations*. P.A. Facione.
- Hao, Z., Cao, J., Li, R., Yu, J., Liu, Z., & Zhang, Y. (2026). Mapping student-AI interaction dynamics in multi-agent learning environments: Supporting personalized learning and reducing performance gaps. *Computers & Education*, 241, 105472.
- Hwang, G.-J., Lai, C.-L., Liang, J.-C., Chu, H.-C., & Tsai, C.-C. (2018). A long-term experiment to investigate the relationships between high school students' perceptions of mobile learning and peer interaction and higher-order thinking tendencies. *Educational Technology Research and Development*, 66(1), 75–93.
- Murphy, E. (2004). An instrument to support thinking critically about critical thinking in online asynchronous discussions. *Australasian Journal of Educational Technology*, 20(3).
- Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, 220, 105112.