

生成式人工智能在教育剧本杀（Edu-LARP）中的多维测评与差异分析——以高中物理《全反射》为例

Multidimensional Assessment and Differential Analysis of Generative AI in Educational Live Action Role-Playing (Edu-LARP): A Case Study of "Total Internal Reflection" in High School Physics

潘妍蓉

复旦大学高等教育研究所

[*panyr24@m.fudan.edu.cn](mailto:panyr24@m.fudan.edu.cn)

【摘要】 针对游戏化学习中沉浸感与知识建构的张力，本研究构建了生成式 AI 作为 NPC 的物理教育剧本杀模式。以《全反射》为例，采用“模拟学习者”法，从概念建构、支架适应性及叙事融合三维度对比测评 ChatGPT 与豆包。结果表明：豆包偏向“规则导向建构”与“程序化硬支架”，能有效降低新手认知负荷；ChatGPT 擅长“直觉导向推演”与“诊断性软支架”，利于促进高阶思维。同时，两者均易诱发“角色冲突”从而破坏沉浸感。本研究明晰了 AI 在游戏化教学的应用边界，为优化人机协同探究课堂提供了实证支撑。

【关键词】 生成式人工智能；物理教育剧本杀；多维测评；人机协同

Abstract: Addressing the tension between immersion and knowledge construction in game-based learning, this study constructs a physics Edu-LARP (Educational Live Action Role-Playing) model utilizing Generative AI as non-player characters (NPCs). Using "Total Internal Reflection" as a case study and employing the "Simulated Learner" method, ChatGPT and Doubao were evaluated across three dimensions: conceptual construction, scaffolding adaptability, and narrative integration. The findings reveal that Doubao favors "rule-directed construction" and "procedural hard scaffolding," reducing cognitive load for novices. ChatGPT excels in "intuition-directed deduction" and "diagnostic soft scaffolding," which facilitates higher-order thinking. However, both models are prone to inducing "role conflicts," thereby undermining immersion. This research clarifies the application boundaries of AI in gamified instruction, providing empirical support for optimizing human-AI collaborative inquiry-based classrooms.

Keywords: Generative AI, Edu-LARP in physics, Multidimensional Assessment, Human-AI Collaboration

1. 前言

1.1. 研究背景

游戏化学习（Game-based Learning）作为一种创新的教学范式正受到学界的广泛关注。基于建构主义学习理论(Constructivism) (Piaget, 1936; Vygotsky, 1978)，游戏化环境提供的认知情境，有助于其通过同化与顺应自主构建知识框架，从而深化对学科知识的理解与迁移。然而，如何在游戏化教学中平衡“沉浸式体验”与“深度知识建构”仍是该领域面临的核心挑战。

随着生成式人工智能(Generative AI)技术的飞速发展,其作为智能认知工具介入教育场域,为强化课堂深度、优化教学质量提供了新的契机。特别是将生成式AI引入教育剧本杀(Edu-LARP) (Deterding, 2018)这一新兴教学模式中,探究其在提供教学支架(Instructional Scaffolding) (Wood, Bruner & Ross, 1976)及扮演非玩家角色(NPC)方面呈现何种特征与潜能,是推动人工智能赋能教育改革的重要前提。

1.2. 研究目的

本文以“剧本杀融入物理课堂”为例,构建人机协同的游戏化教学新生态,对比测评。

ChatGPT-5.2 Thinking 和豆包-深度思考模式在该教学场景下的三层维度,旨在明确不同生成式 AI 模型在游戏化课堂中的适用边界与应用策略。

2.研究过程

2.1. 情境设计

本研究依托高中物理教材选择性必修课程中关于《全反射》的教学单元,设计了物理课堂教育剧本杀(Edu-LARP)课例。该课例创设了高能激光安防系统失效致使稀世钻石“海之眼”失窃的沉浸式悬疑情境。研究选取探案过程中两个探究片段作为案例切片,如表 1 所示。

表 1 生成式AI 角色设定与教学引导提示词设计

角色指令与引导策略	
案例切片一	你现在是“深蓝博物馆案件的物理顾问AI”。学生正在审讯嫌疑人技工。技工声称在水下 20 米看到了 50 米外岸上的人。引导学生利用折射定律和临界角计算视线角度,让学生自己发现证词中的物理漏洞。
案例切片二	你现在是“刑侦鉴识科的物理专家”。学生怀疑这根弯曲的玻璃棒是作案工具。已知棒的折射率为 1.50,水的折射率为 1.33。请引导学生建立几何模型,推导:为了保证激光不泄露,玻璃棒的最小弯曲半径应该满足什么条件?

2.2. 实验交互

根据布鲁姆在认知领域的教育目标分类学(Anderson & Krathwohl, 2001),本研究将处于“记忆与理解”水平的学生称之为“新手学生”,将处于“应用与分析”水平的学生称之为“中等学生”,将处于“评价与创造”水平的学生称之为“专家学生”。结合物理学科的特点,三类学生的认知分别呈现以下特点:

表 2 基于布鲁姆分类学的模拟学习者认知特征描述

物理认知行为与思维特征	
新手学生	处于知识的初步习得阶段,能够回忆基础物理概念,但在处理跨情境问题时存在迷思概念(Misconception),难以建立新旧知识间的有效连接,常表现为直觉性迷思与记忆混淆。
中等学生	学生具备基本的应用能力,能识别常规物理情境并尝试运用定理解决问题。但在面对复杂问题情境时,其模型建构能力(Model Construction)尚不成熟,常因忽略边界条件等映射错误而导致模型失效的现象。

专家学生 学生具备高阶思维能力，能够批判性地审视问题预设条件，并基于物理第一性原理提出独创性假设。其不仅能熟练进行复杂逻辑推演，更展现出极强的元认知监控能力(Metacognitive Monitoring)。

在交互阶段，由具备物理学科背景的研究者扮演“模拟学习者”，以新手、中等、专家三种学生的认知特点与 ChatGPT-5.2 Thinking 和豆包-深度思考模式进行测试。本研究共收集了两个 AI 模型在三个预设探究情境下的 6 份完整交互记录，总计约 8000 字中文字符。

2.3. 评估设计

为确保评估结论的科学性，本研究对文本数据进行了人机协同交互的编码分析(Coding Analysis)。自下而上地提取关键对话节点，编码结果如表 3 所示：

表 3 人机协同交互文本编码表

评价维度	编码标签	判别标准	原始文本提取示例
物理概念准确性与逻辑性	规则导向型建构	严格检索教科书定义，倾向将物理情境直接抽象为数学符号与标准公式表达。	豆包：“在计算视线角度前，我首先要明确折射定律的使用条件及公式表述...”
	直觉导向型推演	运用第一性原理与形象化类比，能定性解释非理想或复杂边界形弯曲结构想象成对光线的‘挤压’条件下的物理机制。	ChatGPT：“你可以把这里的锥形作用...”
教学支架类型与适应性	程序化硬支架	将复杂建模任务拆为单一计算步骤，直接下达计算指令或要求代入公式。	豆包：“折射定律是 $n_1\sin\theta_1 = n_2\sin\theta_2$ ，在这里， θ_1 代表的是...”
	诊断性软支架	不直接纠错，而通过逻辑反问或揭示几何矛盾点来引发学生的认知冲突。	ChatGPT：“如果按这个角度推导，岸上的人应在 8 米处而非 50 米处，这就产生了矛盾...”
角色沉浸感与叙事融合	角色冲突	脱离设定的 NPC 身份，使用明显的“教师”或“客服”语用特征，破坏悬疑情境。	两款模型均有出现：“同学，请注意这里的全反射临界角条件...”
	角色一致性	物理释疑话语自然嵌入探案悬疑逻辑，口吻符合设定的专家或顾问人设。	ChatGPT：“探员，供词在物理学上存在致命漏洞，让我们来还原一下案发时的光路...”

3.结果与讨论

3.1. 物理概念建构的准确性与逻辑性

依据 Kortemeyer (2023) 的评估标准与专家-新手认知理论 (Chietal,1981)，研究发现两款模型在处理非常规物理问题时展现出显著差异：豆包倾向于“规则导向型建构”，表现出典型的新手特征。其逻辑推演高度依赖表面特征与方程式检索。ChatGPT 则展现出“直觉导向型推演”的专家特质。它能运用第一性原理建构问题的心智模型，具备更敏锐的物理直觉，逻辑推演过程更为严密。

3.2. 教学支架能力的风格差异：硬支架与软支架

参考 Saye 和 Brush(2002) 的理论,分析发现两款模型代表了截然不同的支架范式,对学习者最近发展区(ZPD)的响应各异:豆包侧重程序化的“硬支架”:它倾向于将复杂推理拆解为线性的步骤引导。这种方式虽能显著降低新手的认知负荷,但也因过度提示(Over-scaffolding)剥夺了学生自主建模的机会。ChatGPT 擅长诊断性的“软支架”:它通过构建“认知冲突”与反证法引导学生自省,实现深度学习。然而,这种即时性支架对基础薄弱者可能造成心理挫败感。

3.3. 角色一致性、沉浸感与角色冲突

基于 Edu-LARP 框架 (Deterding, 2018) 与角色理论 (Mead, 1934), 话语分析显示出人机协同中的角色困境:两款模型普遍存在“好为人师”的倾向,难以完全剥离“AI 助手”的身份标签。这种说教模式违背了玩家在探案情境中对 NPC 的角色期待,构成了典型的“角色冲突”。

4. 结语

4.1. 结论与建议

在人机协同的游戏化教学中,不存在绝对完美的单一模型。教师需根据学生的认知层级进行动态使用 AI 模型,并强化提示词中的角色约束,有效平衡“知识建构”与“沉浸体验”。

基于此,本研究提出以下三点实践建议:动态使用模型:针对新手学生,教师应优先调用豆包进行结构化引导;面对中高阶学生,则动态切换至 ChatGPT 激发其对物理原理的深度反思。设定延迟反馈机制:为防范 AI“过度支架”剥夺学生建构机会,教师需设定提示词“沉默期”。要求 AI 在直接给出结论前,先引导学生自主推导,从而明确人机协作边界。精细化角色提示设定:针对 AI 易出现的“教师说教”倾向,教师在应用时需设计角色扮演提示词,规避生硬的教学话语,确保物理释疑自然融入案件叙事,维护玩家的心流体验。

4.2. 局限与展望

本研究主要基于“模拟学习者”的理想化实验环境,缺乏真实课堂中复杂交互噪音的干扰数据。未来的研究将致力于将该 Edu-LARP 系统部署到真实的高中物理课堂中。此外,如何利用 AI 的 API 接口实现“动态难度调节(DDA)”,根据学生的回答实时切换 AI 的支架风格,将是实现真正个性化游戏化教学的关键突破口。

参考文献

- 中华人民共和国教育部(2020)。普通高中物理课程标准(2017 年版2020 年修订)。人民教育出版社。
- 王思悠和朱广天(2025)。剧本杀游戏融入高中物理课堂的教学模式初探——以《牛顿庄园奇案》为例。湖南中学物理, 3, 25-30。
- 方士心、陆一和盛煜森(2025)。面向人才选育的高考难度理论分析:思维品质优先的失分概率。华东师范大学学报(教育科学版),43(09),54-68。
- Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development*, 50(3), 77–96. <https://doi.org/10.1007/bf02505026>