

人工智能在高中信息技术教学中的应用评估：以豆包、通义千问和文心一言为例

Evaluation of the Application of Artificial Intelligence in High School Information Technology Education: A Case Study of Doubao, Tongyi Qianwen, and Wenxin Yiyan

王子轩¹, 杨宁^{2*}

^{1, 2} 福建师范大学教育学院

* 772416452@qq.com

【摘要】 本生成式人工智能的到来为课堂教学注入的新的动力，尤其是在高中信息技术的教育中展现了巨大的潜力，本研究分析了豆包、通义千问、文心一言三个 AI 平台在高考信息技术中的解答表现，为学生和教育工作者选择 AI 平台提供了依据。

【关键词】 生成式人工智能；高中信息技术；AI 平台

Abstract: The arrival of generative artificial intelligence has injected new momentum into classroom teaching, especially demonstrating significant potential in high school information technology education. This study analyzes the performance of three AI platforms—Doubao, Tongyi Qianwen, and Wenxin Yiyan—in answering questions related to the college entrance examination (Gaokao) information technology subject, providing a basis for students and educators to choose AI platforms.

Keywords: keyword one, keGenerative Artificial Intelligence; High School Information Technology; AI Platforms

1. 前言

ChatGPT 的发布标志着生成式 AI 进入教育领域。本研究聚焦 AI 平台在高考信息技术教学中的应用潜力[1]。核心研究问题：1) AI 平台可提供哪些教学支持？2) 各平台性能差异及适用性？通过对比主流平台的解题准确性、逻辑性与解析深度，为教学实践提供参考经验。

2. 相关研究

国内外学者对大语言模型在教育中的应用方式以及大语言模型在知识测验中的表现均有不同程度的研究。隋小波在研究中指出，AI 技术能够显著激发学生的学习兴趣，引导学生进入深度学习状态[2]。廖栩通过“AI 双师课堂”的教学案例，探讨了人工智能技术在信息技术教学中的实际应用及其效果[3]。研究表明，“AI 双师课堂”能够个性化地支持学生的学习需求，提升学习兴趣，并优化教学过程。

在关于 AI 工具的选用上，Olesya M. Tolstykh 和 Tamara Oshchepkova 系统分析了 150 种可供语言教育使用的 AI 工具，并提出了针对英语教学的分类方法。文章指出，这些工具能够在内容创建、评估、辅导和课程规划中为教师和学生提供支持，同时也强调了 AI 工具在教育中的潜在限制，如隐私问题和认知过载。研究为语言教师如何高效选择和应用 AI 工具提供了实用指南[4]。

尽管当前已有多项研究表明，AI 在教育领域，特别是在高考辅导中的潜力巨大，但仍存在一些研究空白。首先，大部分研究关注 AI 在基础教育中的应用，针对高中信息技术学科的研究仍然较为匮乏。其次，现有的评估标准过于单一，主要集中在准确性和完整性上，较少

涉及平台的逻辑性和解析深度，这一点在复杂题目和推理题目的解答中尤为重要。因此，如何提升 AI 在逻辑推理题目中的表现，并进一步提升平台解析深度，是未来研究的重点。

3. 研究设计

3.1. 研究目的

本研究的主要目标是对比多种主流的 AI 平台在解答高考信息技术类题目时的表现，分析它们在准确性、逻辑性、解析深度、语言表达等维度上的优势与不足，并为学生和教育工作者提供相关参考，以便在备考过程中选择最合适的 AI 工具。

3.2. 研究问题

本研究的主要目的是对比多个 AI 平台在解答高考信息技术类题目时的表现，基于此，提出两个本研究的核心研究问题：1、AI 平台在哪些方面可以提供支持？2、哪个平台的更适合此类教学，其性能如何？

研究将比较不同 AI 平台对高考信息技术题目的解答准确性，评估哪些平台能提供最贴合标准答案的解答，并进一步分析其解答的准确率。

本研究采用实证分析方法，结合定量分析，对比多个 AI 平台在解答高考信息技术题目时的表现。

3.3. 前期工作

本研究将选取 2023 年和 2024 年浙江省“技术”高考选科中的信息技术部分真题的选择题部分。将这些题目输入到各个 AI 平台（如豆包、文心一言、通义千问）进行解答。

对每个 AI 平台给出的答案进行量化评分，依据准确性、解析深度、逻辑性、语言表达等标准进行评分，并计算出各个平台的综合评分。使用统计方法对不同平台的表现进行比较，判断它们在不同维度上的差异是否具有统计学意义。

为了保证研究的科学性和代表性，本研究将从收集的数据中选取一组具有代表性的 5 道高考信息技术选择题作为样本，确保难度适中，避免极端简单或极端复杂的题目，能够真实反映学生的解题水平和各平台的表现。为了确保研究结果的可比性和科学性，本研究将为每个 AI 平台的答案设计详细的评分标准。

表 1
评分标准表

标准\得分	0	1	2	3
回答准确性	答案错误	回答部分错误	回答有小部分不足	答案正确
解析步骤	没有提供任何解析步骤	解析步骤不完整，仅覆盖部分要点	解析步骤基本完整，但缺乏细节	解析步骤详细完整，逻辑清晰
解析质量	没有任何解题思路或解析	提供了解题思路，但逻辑不清晰	解题思路清晰，但解析有细节缺失	提供完整、清晰且详细的解题思路和解析

4. 数据分析

4.1. 数据整理与统计

在数据收集完成后，首先对收集到的数据进行整理和统计，后续进行分析和可视化。

将每个平台的回答按照题目编号、题目类型（选择题、非选择题）和评分维度（准确性、解析步骤、解析质量）分类整理。量化评分上，统一用分值（0-3 分）记录每个平台的表现，确保数据结构清晰。

4.2. 数据预处理

检查评分数据是否存在异常值（如错误输入或漏评分），对无效数据进行清理。分类整理选择题和非选择题的数据，便于分组对比。

4.3. 数据计算

在本研究中，为了全面、客观地评估各 AI 平台在高考信息技术选择题解答中的表现，采用了“加权综合评分法”对数据进行分析。该方法通过对不同评分维度（准确性、解析步骤、解析质量）赋予不同权重，将每个平台的多维度评分归一化为一个综合得分，便于平台间的比较和排名。准确性：平台提供的答案是否正确，权重为 50%，这一维度在解答选择题中最为重要，直接反映平台的基本能力。解析步骤：平台是否提供了逻辑清晰、完整的解题步骤，权重为 25%，该维度体现平台对解题过程的支持。解析质量：平台提供的解析是否详尽且易于学生理解，权重为 25%，这一维度评估平台在教学辅助中的能力。

综合得分的计算公式为：

综合得分 = 准确性得分 * 50% + 解析步骤得分 * 25% + 解析质量得分 * 25%

该公式的设计基于权重的加权平均计算方法，将三个评分维度的重要性差异量化为权重。

4.4. 结论

4.4.1. 选择题分析

豆包：综合得分 2.50（准确性 2.57/3），解析步骤与质量均达 2.43 分，显著优于其他平台。

通义千问：综合得分 2.18，解析质量（2.14）与步骤（2.29）存在逻辑断层。

文心一言：解析质量最低（2.00），综合得分 2.11，教学支持能力较弱。

4.4.2. 非选择题分析

各平台综合得分趋近（豆包 1.58，通义千问 1.50，文心一言 1.46），普遍存在步骤缺失与解析浅表化问题。豆包在步骤清晰度（1.83）上相对占优，但均未达到教学辅导深度需求。

文心一言不仅在准确性上没有明显的优势，而且在解答过程中，缺乏足够的教学辅助和解题步骤的详细解析。其低分表明，文心一言在解答非选择题时的表现不如其他平台，可能无法提供足够的教学支持和思路引导。

5. 讨论

5.1. 对现有 AI 平台的分析和反思

准确性作为 AI 平台核心竞争力存在显著差异：豆包在高考信息技术试题中表现最优，但在高难度选择题和非选择题仍存在不足；通义千问与文心一言在处理复杂问题时错误率较高。这表明现有平台在专业领域知识深度理解方面仍需加强，需提升细化解题与精准生成答案的能力。

解析逻辑性呈现明显梯度：豆包能清晰展示解题过程，得分显著高于另两个平台；通义千问步骤常缺乏连贯性，文心一言解析过于简略。建议平台应强化解题逻辑的系统性构建，确保步骤衔接合理、可操作性强，有效支持学生理解。

在解析深度方面，豆包对复杂题目的细节阐述仍有提升空间，而通义千问和文心一言普遍存在解析深度不足的问题，影响教学支持效果。未来需重点优化关键节点的深度解析，特别是在高难度题目中加强思路展开。

5.2. 研究的局限性与未来研究方向

5.2.1. 研究的局限性

本研究仅基于浙江省高考信息技术选择题的数据分析，未来研究可以扩展到其他方面，以更全面地评估 AI 平台的教学能力。数据来源局限于部分高考真题，样本数量相对较少，可能对研究结果的普适性造成一定影响。

5.2.2. 未来研究方向

扩展样本范围，纳入更多学科和题型的数据进行分析。引入学生用户体验的调查数据，结合定性分析，进一步研究 AI 平台的交互设计和使用效果。探讨 AI 平台在个性化教育的潜力，为 AI 技术在教育领域的创新应用提供更多实证依据。

5.2.3. 总结

本研究基于浙江省高考信息技术的定量分析，对豆包、通义千问、文心一言三个 AI 平台的解答能力进行了系统评估。豆包在综合能力上最优，因此更加适合信息技术学科的教学，但在处理较难或较为复杂的问题上仍然表现不佳，例如较难的选择题和非选择题等。随着 AI 技术的不断发展，其在教育领域的应用将更为广泛，未来的研究与实践应进一步聚焦于提升平台的教学质量和学习体验，为教育公平和个性化学习创造更大价值。

参考文献

- 熊明辉,池骁. 论生成式大语言模型应用的安全性:以 ChatGPT 为例[J]. 山东社会科学,2023(5):79-90.
- 隋小波. AI 促使深度学习的信息技术教学探究——以“Python while 循环”项目设计为例[A] 北京大学出版社 2023 年教育数字化转型与智能教育发展研讨会论文集[C]. 北京大学出版社, 北京未名智慧教育科技有限公司, 2023: 4.
- 廖栩, 陈雪松. 人工智能在信息技术课堂教学中的应用探究——以 AI 双师课堂《救援呱呱》为例[J]. 中国现代教育装备, 2020, (08): 25-27.
- Tolstykh O. M., Oshchepkova T. Beyond ChatGPT: Roles that artificial intelligence tools can play in an English language classroom[J]. Language Teaching and Learning, 2023(5): 58-75.