

# 大学生文本 vs 生成式人工智能文本：相似度与学生评价

## University Student Output vs Generative AI Output: Similarity and Student Evaluation of Generative AI Output

周晓华<sup>1\*</sup>, 柴文玉<sup>2</sup>, 林敬新<sup>1</sup>

<sup>1</sup> 岭南大学, 教与学中心

<sup>2</sup> 岭南大学, 研究生院

\*amyzhou@ln.edu.hk

**【摘要】** 本研究通过分析 33 名大学生的文本与 GenAI 生成文本的相似度（分为高、低相似组），结合学生评价，探究影响相似度的可能原因。研究发现：一、两组均认为 GenAI 回答具有参考价值，结构清晰、内容准确、符合原文，但缺乏深度和细节支撑；二、高相似组更关注其是否帮助理解阅读材料和完成学习任务，尽管意识到 GenAI 回答缺乏深度但自身回答仍与其更相似；三、低相似组更重视其全面性及其对思考的启发，并认为在独立思考和延伸拓展方面 GenAI 无法代替人类阅读。研究启示高校教师可通过培养学生深度学习导向，设计需独立思考和拓展创新的学习活动及评价方式，降低学生对 GenAI 的依赖，发展学生思维的独特性和创新性。

**【关键词】** 生成式人工智能；大学生；相似度

**Abstract:** This study examined the text similarity between outputs produced by university students and those by GenAI (categorized into high- and low-similarity groups) in completing learning tasks, along with students' evaluations of GenAI output, to identify possible reasons for the observed similarity. A total of 33 university students in Hong Kong participated in the study. The findings reveal that: 1) Both groups consider GenAI output to be valuable references, characterized by clear structure, accurate content, and alignment with reading material, yet lacking depth and supporting details; 2) The high-similarity group focuses more on whether GenAI output helps with understanding reading materials and completing learning tasks, and despite recognizing the lack of depth in GenAI output, their own output remain more similar to GenAI output; and 3) The low-similarity group places greater emphasis on the comprehensiveness of GenAI output and their help with stimulating thinking, and maintains that GenAI cannot replace human reading in terms of independent thinking and extended exploration. University instructors should foster a deep learning orientation among students, design learning activities and assessments that require independent thinking and innovation, thereby reducing students' reliance on GenAI and cultivating the uniqueness and creativity of their thinking.

**Keywords:** Generative AI; university student; similarity

## 1. 引言

近年来，大学生在学习中使用人工智能(AI)越来越普遍。一项国际大型调查显示大约 86% 的大学生在 学习 中使用 AI (Digital Education Council, 2024)。在中国，超过 70% 的大 学生 使用 生成式人工智能(GenAI)，主要用于生成文本和提取信息 (Xiao et al., 2025)。伴随着愈发普遍的使用，教育界对大学生过度依赖 GenAI 的担忧逐渐产生。GenAI 能够快速生成解答，提供“认知捷径” (Zhai et al., 2024)，同时其使用能提升大学生解决问题的信心和效率，这些都可能促使大学生更加依赖 GenAI (Zhang & Xu, 2024)。

尽管使用 GenAI 可以帮助学生高效完成学习任务，过度依赖 GenAI 则会阻碍大学生的独立学习（Zakaria et al., 2025），导致自我调节学习中的“元认知懒惰”（Fan et al., 2025）。考虑到愈发广泛的使用和过度依赖产生的负面影响，厘清哪些因素可能引发大学生对 GenAI 的依赖变得至关重要。已有研究显示，除了社会人口信息（例如年级、专业）和 AI 素养外，学生批判性思考能力以及对 GenAI 的信任也会影响他们对 GenAI 的依赖（Ding 等，2023；Hou 等，2025）。Hou 等的研究（2025）发现批判性思维与大学生合作性地、反思性地以及谨慎地使用 GenAI 显著正相关，主动进行批判性思考的大学生更有可能对 GenAI 回答进行批判性评估。此外，对 GenAI 的信任程度也影响大学生对它的依赖。Ding 等的研究（2023）发现尽管 ChatGPT 在回答问题时给出了不准确的答案，大部分大学生仍然信任它可以给出正确答案，甚至当学生被明确告知 AI 可能产生错误时，近一半的学生仍然相信它的答案。

基于上述背景，本研究试图从大学生在完成课业任务时的回答与 GenAI 回答之间的文本相似度这一角度来探究哪些因素可能与大学生依赖 GenAI 有关。

## 2. 研究设计

### 2.1. 研究背景和研究对象

本研究从香港一所高校的一门社会科学专业授课制硕士研究生课程中招募研究对象。该课程每周一次研讨课，持续五周。在每周研讨课之前，学生需要阅读一篇文章并回答三道阅读问题，帮助学生理解研讨课的主题，更好地参与研讨课的课堂讨论。本研究持续四周（第二-五周），期间学生需每周在线回答与阅读材料相关的问题。同时，由 GenAI（ChatGPT4.0）对阅读问题生成的回答也作为参考提供给学生。此外，学生需分享他们对 GenAI 回答的看法：“您认为 GenAI 生成的关于问题一至问题三的回答如何？”

### 2.2. 数据分析

本研究首先使用 R（*quanteda* 包）分析了每位学生与 GenAI 回答的余弦相似度（绝对值在 0 与 1 之间，数值越大代表越相似），然后针对学生对 GenAI 回答的评语用 Nvivo 进行主题编码分析，以探索学生对 GenAI 回答的评价与其回答相似度之间的联系。

## 3. 研究结果

共 33 位大学生在四周中完成了 61 份有效作答（包含 3 道阅读问题和对 GenAI 回答的评语）。依据学生回答与 GenAI 回答的相似度，将其分为高相似组（ $n=30$ ，相似度 0.399-0.763，表示学生回答与 GenAI 回答相似度较高）和低相似组（ $n=30$ ，相似度 0.129-0.373，表示学生回答与 GenAI 回答相似度较低），剩余 1 份（相似度 0.375）未纳入分析。

### 3.1. 总体评价

在 60 条评语中，47 条(78%) 包含正面和负面评价，13 条(22%) 仅包含正面评价。在高相似组和低相似组中，总体来说 GenAI 回答都被认为具有参考价值。在高相似组中，GenAI 回答不仅被认为有参考价值，也被视为“草稿”。案例请见表 1。

### 3.2. GenAI 回答的优点

高相似组和低相似组均提到的 GenAI 回答的优点包括：结构清晰、内容准确、符合原文等。两组的主要区别在于：低相似组更在意是否全面（高相似组 5 处 vs 低相似 9 处）；高相似组更看重是否有效总结了原文的核心观点（高相似组 9 处 vs 低相似组 4 处）以及是否包含正反两方面的评价（高相似组 8 处 vs 低相似组 3 处）。

### 3.3. GenAI 回答的缺点

高相似组和低相似组均认为 GenAI 回答缺乏深度和支持性细节。有趣的是，高相似组更频繁指出 GenAI 回答缺乏深度（高相似组 16 处 vs 低相似组 5 处），但其自身回答仍与 GenAI 回答更相似。以下为一组高相似组案例（T3\_24，相似度 = 0.406）：

- GenAI 回答：“.....此外，尽管文章批判了现有理论，但仍可提出更具体的替代性理论模型或框架，以指导未来的研究.....”
- 学生对 GenAI 回答的评语：“我的主要担心是，对理论贡献的批评略显单薄——它指出文章应当提出替代模型，但并未具体说明可以用哪种模型。”。
- 学生回答：“.....尽管它指出现有理论的缺陷，但并未提出一个完整详尽的新模型.....”

### 3.4. 人机协作

|                | 高相似组                                                                               | 低相似组                                                                                                                                                         |
|----------------|------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 对              | “总的来说，它们做初稿很优秀，忠实地呈现了原文，为进一步精炼或直接使用提供了扎实的基础，前提是用户已对照原文核实了其准确性。” (T2_3)             | “总的来说，生成式人工智能的回答确实是一个很好的参考，既为我们提供了一些思路，也补充了我们自己的想法。” (T2_26)                                                                                                 |
| GenAI 回答的      | 总体评价 (T4_8)                                                                        |                                                                                                                                                              |
| GenAI 回答的优点    | “针对问题一，GenAI 能够有效地从摘要和引言中概括出关键要点，使我们能够基本快速了解文章内容。” (T4_1)                          | “全面性：回答涵盖了研究对象、驱动因素、发展动态及结论等关键部分，未遗漏核心要点。” (T2_16)                                                                                                           |
|                | “对问题三，我认为回答是平衡的，兼顾了优点和不足。优点表述清晰：……提出的不足也合理：……。” (T4_15)                            |                                                                                                                                                              |
| GenAI 回答的缺点    | “这个回答感觉有点太停留在描述层面了。它报告了文章的内容，但本可以更深入探讨更广泛的影响，如跨国节关注上有所不足，人工智能未从原文中提取实证证据。” (T2_40) | “不过，这些回答有时缺乏具体性，只提供了一个概括性陈述而缺少实例。在细                                                                                                                          |
|                | 或者将如何塑造中国长期的社会结构。” (T4_1)                                                          |                                                                                                                                                              |
| GenAI 辅助阅读     | “人工智能帮助我精确把握阅读材料中的核心要点，并简要阐释文本中较为抽象的概念。” (T5_6)                                    | “结构上，它帮助我组织观点和文献中的实证证据，同时提供给我一个思考范式。” (T2_23)                                                                                                                |
| GenAI 无法替代人类阅读 | “批判性地阅读原文对用户验证其准确性来说仍是至关重要的……” (T3_33)                                             | “最好自己阅读全文。尽管翻译和理解会花费更长时间，但可以学到更多东西，促进思考。” (T2_19);                                                                                                           |
|                |                                                                                    | “AI 忠实地复现了原文的局限性，而没有突破这些边界。此处需要人为干预：我们应该明确指出“仅以一个中外合办大学+一个英国硕士项目作为研究样本”属于结构上的局限，不仅是方法论上的脚注，而且要有一个后续研究调查例如非洲或者中亚地区的毕业生，考察当种族因素和非西方高校纳入讨论后，研究结果是否仍成立。” (T4_10) |

在人机协作方面，学生评语中出现了两个主题：“GenAI 辅助阅读”和“GenAI 无法替代人类阅读”。在“GenAI 辅助阅读”方面，高相似组更认可 GenAI 在理解原文内容上的帮助，低相似组则更看重其对思考上的启发。在“GenAI 无法替代人类阅读”方面，高相似组基于准确性层面，低相似组则基于思考和拓展层面提出该观点。

表 1 高相似组和低相似组编码案例

## 4. 讨论与建议

### 4.1. 批判性评价不足以帮助学生产生不同于 GenAI 的回答

高相似组更频繁地指出 GenAI 回答缺乏深度、流于表面，但其自身回答与 GenAI 的相似度却较高。这一矛盾现象表明：部分学生虽能意识到 GenAI 回答的局限性，却可能受到自身

认知、表达能力限制，难以产出不同于 GenAI 的更有深度的回答。因此，教师需针对学生的困难及原因（如英文写作能力有限）帮助已意识到 GenAI 局限的学生形成更完善的回答。

#### 4.2. 深度学习可能减少学生对 GenAI 回答的依赖

本研究发现在高相似组中学生展现出浅层学习的倾向（例如更关注 GenAI 对完成学习任务的帮助），而在低相似组中学生展现出深度学习的倾向（例如宁愿多花时间阅读原文独立思考）。当学生更想要促进思考而不是高效完成任务时，其与 GenAI 回答的相似度更低，更会产出不同于 GenAI 的回答。这说明，学生对 GenAI 回答的依赖不仅是一个新出现的与技术有关的问题，更是一个长久的教育问题：如何鼓励学生深入思考，而不仅是完成学习任务。尤其在 GenAI 日新月异的时代，面对更加快速高效的“认知捷径”，在教学中如何鼓励学生花时间去自己动脑、付出努力得到答案、提高自身能力尤为重要。

#### 4.3. 人机协作中人类的不可替代性

低相似组比高相似组更频繁指出“GenAI 无法代替人类阅读”，主要体现在准确性、思考和拓展层面。该发现不仅呼应了已有研究中关于对 GenAI 的信任影响大学生对其依赖的结论（如 Hou 等，2025），也凸显了独立思考和拓展创新的不可替代性。这启示高校教师可设计更强调独特性和创新性的学习任务和评估方式，培养大学生独立思考和积极创新。

### 参考文献

- Digital Education Council. (2024, August 2). *Digital education council global AI student survey 2024*. <https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024>
- Ding, L., Li, T., Jiang, S., & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, 20(1), 63.
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., ... & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489-530.
- Hou, C., Zhu, G., & Sudarshan, V. (2025). The role of critical thinking on undergraduates' reliance behaviours on generative AI in problem-solving. *British Journal of Educational Technology*, 56(5), 1919-1941.
- Xiao, L., Pyng, H. S., Ayub, A. F. M., Zhu, Z., Gao, J., & Qing, Z. (2025). University students' usage of generative artificial intelligence for sustainability: A cross-sectional survey from China. *Sustainability*, 17(8), 3541.
- Zakaria, N. Y. K., Hashim, H., & Jamaludin, K. J. (2025). Exploring the impact of AI on critical thinking development in ESL: A systematic literature review. *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence*, 330-347.
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1), 28.
- Zhang, L., & Xu, J. (2025). The paradox of self-efficacy and technological dependence: Unraveling generative AI's impact on university students' task completion. *The Internet and Higher Education*, 65, 100978.