Proceedings of the 5th International Conference on International Law and Legal Policy
DOI: 10.54254/2753-7048/2026.38468

Generative AI Use and Critical Thinking among University Students

Yi Yan[1], Chentao Zhu[2,*]

[1]Hanvos Kent High School, Ningbo, China
[2]School of Foreign Languages, Taizhou University, Taizhou, China
**Corresponding Author. Email: jimomega2113@gmail.com*

Abstract. How students use artificial intelligence and what they learn from it cannot be ignored. According to how often people use AI tools, the groups were divided, and then descriptive statistics, Welch's ANOVA test and independent-sample t-tests were carried out to compare the critical-thinking scores of these groups. Based on the above description, the mean critical-thinking scores of the four groups using AI are relatively high. However, Welch's ANOVA showed that there were no significant differences among the four groups ($F(3, 24.71) = 1.586$, $p = .218$). The two groups were divided into regular and non-regular AI users; the mean critical-thinking scores of the regular-use group (50.06) and the non-regular-use group (45.42) were significantly higher for the regular-use group; $t(100) = 2.948$, $p = .004$, and Cohen's $d = .750$. Based on the above results, the impact of AI on students' critical thinking ability is not straightforward; it may vary according to how they use AI. AI can help students learn better by learning to think independently; however, if they rely on AI too much, they may not develop their own thinking skills. Given the small sample size, uneven distribution of groups, self-reported data, cross-sectional design, etc., only associations can be concluded from the results and not cause-and-effect relationships.

Keywords: Artificial intelligence, AI application behaviour, critical thinking, cognitive offloading, students and learners

1. Introduction

Generative Artificial Intelligence (AI) has been applied in all aspects of students' study, including information collection, idea generation, problem-solving and homework help. ChatGPT and other tools offer quick answers and different viewpoints, but people are also beginning to wonder how reliable this information is and whether they will still verify it themselves. Research on critical thinking has shown that it is necessary to weigh evidence, determine assumptions, detect flaws in logic, and draw reasonable conclusions [1, 2]. The reasons for AI-assisted learning are that the AI-generated answers may be incorrect, incomplete or misleading.

There is no single consensus that AI applications are good or bad. One student may ask AI for the full answer, and another can use the same tool to get different explanations, test an idea, or find feedback before deciding. Therefore, this paper will compare the critical-thinking scores of different

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.38468

patterns of AI use. It does not assume that a high frequency is necessarily related to strong or weak thinking ability.

2. Literature review

Digital technology is not inherently good or bad for critical thinking. Moustaghfir and Brigui think that digital information should be judged critically and not blindly accepted [1]. The same problems have been reported for social media; that is, the speed and selection of information dissemination can alter people's perceptions [3]. The discussion has now moved from the spread of technology to what students do with the information they have gained.

Generative AI has made this problem more obvious; it can take over some of the work that students would have to do themselves. Larson and others think that these tools can change the environment in which students learn and solve problems [4]. Gerlich thinks that, because of cognitive offloading, many people use artificial intelligence tools so often that their own thinking skills have declined [5]. Some other studies are not very pessimistic. Essel and others have pointed out changes in critical, creative and reflective thinking in educational research on ChatGPT [6]. Liu and others have also concluded in their systematic review of English as a Foreign Language (EFL) research that generative AI can help with some aspects of critical thinking, but it also has other problems depending on how it is used in learning [7].

3. Methodology

A Standardised, Structured Questionnaire was distributed to the people. The four areas of the participants' full-time domestic studies were: humanities and social sciences, economics and management, science and engineering, agriculture and medicine, and arts and sports. Responses that were too quick or repeated the same answer were excluded. Finally, 102 good answers were chosen for analysis.

The two parts of the 16 questions are general information and AI usage, as well as critical thinking. The first few columns are grouping variables for the follow-up analysis. According to how often the students used it, the four groups were: non-users (never used), occasional users (1-2 times a week), regular users (3-4 times a week), and intensive users (5 or more times a week). An extra binary question was added to find out how often students use generative artificial intelligence (AI) in their studies, and based on this, the two groups were divided into frequent-users and non-users. The specific classification rules for the two groups should be considered together with the questionnaire coding rules of the original dataset.

The basis of the critical-thinking scale is Ennis's logical structure of critical-thinking skills and Facione's expert consensus report on critical thinking [8, 9]. Some of these have been modified to fit the study context of generative AI, for example, to verify information, check the logic, find problems with bias, and think about the answers that AIs provide. Adaptation aims to link the traditional idea of critical thinking with students' actual application of artificial intelligence, so it should be used as a context-specific questionnaire rather than the original Ennis or Facione measures.

A five-point Likert scale was used, with 1 being "strongly disagree" and 5 being "strongly agree", and the total range of the 13 items was 13 to 65 [10]. A high score means that the person frequently employs critical thinking in AI-assisted study.

The 13 items fall into the following four categories: information verification, logical analysis, bias awareness and deep reflection. A typical case is to verify the information provided by AI with

DOI: 10.54254/2753-7048/2026.38468

textbooks or other reliable sources, find any logical errors, consider different perspectives, and have the AI explain the basis of its conclusion. The aims of the above items are to gauge students' sense of critical thinking, not their actual abilities.

The first part of the questionnaire is to classify students based on how they use AI, and then the second part is the critical-thinking scale for collecting quantitative data. Then, this study compared the critical-thinking scores of all the groups that used AI.

4. Results

4.1. Descriptive results

The first analysis compared the results of the self-reported frequency groups. The mean scores of those who used the service occasionally, regularly and frequently rose to 47.51, 49.62 and 51.03, respectively. The analysis also included an unused category, but its mean score and sample size are not available in the current paper and need to be verified with the original data. Therefore, the increase in all three reported groups is only a pattern in the sample and has not been statistically verified as a trend.

4.2. Welch ANOVA

As the four frequency groups had different numbers of samples, Welch's ANOVA was used for the general test. The test yielded $F(3, 24.71) = 1.586$ and $p = .218$. Since the p-value is greater than 0.05, the groups do not differ significantly according to statistics. The current data cannot show that the critical-thinking scores are different in the four frequency groups.

4.3. Independent-samples t-test

The second division was according to the binary group of the questionnaire item on study-related AI use. The regular-use group had 83 students and the non-regular group had 19 students. The mean scores were 50.06 and 45.42, and the difference between them was 4.64. The independent-samples t-test found $t(100) = 2.948$, $p = .004$. The effect size Cohen's d was 0.750, and the two groups were quite far apart in standard deviation.

Therefore, it can be concluded that the regular-use group had higher self-reported scores for critical thinking in this sample. However, as this is an observational and cross-sectional study, the results can only show an association and do not prove that regular use of AI improves critical thinking ability.

5. Discussion

5.1. Relationship between AI use and critical thinking

In short, the two analyses have reached different conclusions. The mean scores of the three frequency groups were all higher, but the overall Welch's F-test did not show significance. The numbers of the two groups of users were not the same. Therefore, the results do not support the claim that critical thinking has improved continuously with the increase in the use of AI.

Frequency may be too general a measure of how often students use AI. Larson and others have pointed out that generative AI can alter how students learn and think [4]. One student may ask for further explanation or a set of questions and then independently verify the answer. Another may just

DOI: 10.54254/2753-7048/2026.38468

copy the answer and move on. The amount of AI use can be the same, but the degree of independent thinking is significantly different.

Gerlich's talk on cognitive offloading is another way to look at the results [5]. According to this study, many people do not use AI tools because, by doing so, they lose some of their own critical thinking abilities to save effort. This is also why frequency can be misleading. Both students use AI regularly, but one is willing to doubt and verify the results; the other takes the AI's answers as given.

The different current results and those obtained by Gerlich may be due to various reasons, such as different samples, measurement methods, research designs, etc [5]. Gerlich had 666 subjects of all ages and educational backgrounds, and directly studied cognitive offloading. Currently, there are only a few students, and the assessment of self-reported critical-thinking behaviour is based on categories of AI use. Therefore, the two studies should not be considered direct equivalent tests of the same phenomenon.

5.2. AI as a cognitive scaffold

Based on the above results, it can be assumed that AI may function as a cognitive support to help students learn to assess information independently. Liu and others have systematically reviewed generative AI and critical thinking in EFL teaching, and found that when properly applied in class, AI can aid students' learning, thinking and exploration, but excessive dependence on it is also problematic [7]. This view is consistent with the results of the present study that the frequency of AI use alone did not differ significantly among the four groups.

Students will first use AI to obtain explanations, examples, questions and other ideas in practice. Then they can compare the AI-generated information with the textbooks, teachers' explanations, etc. Based on the above information, students can determine which parts are beneficial and revise their own knowledge. At the same time, AI provides the material basis and prompts for thinking; it does not think independently.

However, this explanation should not be given too much weight, as the present questionnaire only gathers students' self-reported behaviour and does not directly observe their interaction with AI. Only the scale items can determine whether students report verification and reflection; it cannot be known how often this behaviour occurs in their studies.

5.3. Comparison with previous studies

The current results are in line with what Essel et al found: they investigated the application of ChatGPT by university students and observed that it affected their abilities in critical thinking, creativity and reflection [6]. Based on the above results, AI-assisted learning can be linked to changes in all sorts of thinking, but the study designs and educational environments are not the same as those in the current survey. Therefore, the present results should be regarded as supplementary rather than direct confirmations.

This paper is also not concerned with the problems of over-reliance on artificial intelligence. Gerlich has shown that if students are too dependent on AI, their own thinking ability will decline as a result [5]. Based on the present study, regular AI users scored higher in critical thinking than non-regular users, but the four-group ANOVA did not achieve statistical significance. Based on the above results, the relationship will vary depending on how AI is utilized and what students actually do with it.

There may be other reasons for the difference. Students who frequently use artificial intelligence may be different from other students in terms of their grades, study methods, motivation, access to

DOI: 10.54254/2753-7048/2026.38468

technology, interest in learning all sorts of things, etc. As these factors were not included in this study, this study cannot attribute the high scores of regular AI users solely to their use of AI.

Therefore, in the future, research should move away from the often-mentioned indices and focus on the actual behaviour of AI, such as checking information, comparing various sources, questioning the reasoning of AI, and correcting one's own ideas. Therefore, researchers will obtain more direct evidence of whether some kinds of interaction with AI are associated with higher levels of critical thinking ability.

6. Conclusion

Therefore, it was decided that there is a difference in the self-reported critical-thinking scores of regular and non-regular AI users. The mean score of the regular-use group was 50.06, and that of the non-regular-use group was 45.42; $t(100) = 2.948$, $p = .004$, and Cohen's $d = .750$. Welch's ANOVA did not find any statistical difference among the four groups of AI use frequency ($F(3, 24.71) = 1.586$, $p = .218$). Therefore, the results do not support the idea that frequent use of AI is directly related to better critical thinking.

The above results show that the effect of AI on critical thinking will differ according to the application mode of AI and, more significantly, by how much students interact with AI-generated content. The previous studies have mentioned the good and bad points of using AI. Generative AI can offer some ideas and help people learn, but if they are too dependent on it, their own thinking will decline.

Some defects have been observed. First, the sample size was small and the groups that used AI were of different sizes. Second, the data are from self-reported questionnaires, and researchers do not know how students actually use AI. Thirdly, a cross-sectional design can only show association at a certain time and not causality. Finally, this paper does not directly measure cognitive offloading or observe how students actually use AI.

In the future, more numerous and widely spread-out samples will be taken to distinguish different kinds of AI applications, and less emphasis will be placed on frequency. Researchers should determine how often students verify the information provided by AI, compare it with other sources, question the AI's reasoning, and revise their own ideas based on the results. Both the long-term observation and the experiment will provide support for the hypothesis that some applications of generative AI affect students' critical thinking.

Author contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Moustaghfir, S., & Brigui, H. (2024). Navigating critical thinking in the digital era: An informative exploration. *International Journal of Linguistics Literature & Translation, 7*(1), 137–143. https://www.researchgate.net/publication/377526220_Navigating_Critical_Thinking_in_the_Digital_Era_An_Informative_Exploration

[2] Critical thinking. (2022, October 12). *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/critical-thinking/

[3] Is social media affecting our critical thinking ability? (2025, November 10). *Stafford Global*. https://www.staffordglobal.org/blog/social-media-critical-thinking/

[4] Larson, B. Z., Moser, C., Caza, A., Muehlfeld, K., & Colombo, L. A. (2024). Critical thinking in the age of generative AI. *Academy of Management Learning & Education, 23*(3), 373–378.

DOI: 10.54254/2753-7048/2026.38468

https://doi.org/10.5465/amle.2024.0338
[5] Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006
[6] Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6*, 100198. https://doi.org/10.1016/j.caeai.2023.100198
[7] Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments, 12*, 48. https://doi.org/10.1186/s40561-025-00406-0
[8] Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 44–48. https://files.ascd.org/staticfiles/ascd/pdf/journals/ed_lead/el_198510_ennis.pdf
[9] Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment*. California Academic Press. https://stearnscenter.gmu.edu/wp-content/uploads/12-The-Delphi-Report-on-Critical-Thinking.pdf
[10] Likert scale. (n.d.). In *Baidu Baike*. Retrieved September 4, 2026, from https://baike.baidu.com/item/Likert scale/2282414