Available on: https://jsshrs.ep-journals.org/ E-ISSN: 3051-3561, P-ISSN: 3051-3553

Journal of Social Science and Human Research Studies

Volume : 02 Issue : 10 October-2026

Exploring the Influence of AI Tools on Critical Thinking Development in English Major Undergraduates

Yanhui Chen

Shanghai Institute of Technology, School of Foreign Languages, Shanghai, China. https://orcid.org/0000-0001-8045-7809

Luoyi Wang

Shanghai Institute of Technology, School of Foreign Languages, Shanghai, China. https://orcid.org/0000-0001-8045-7809

Published Date: October 17, 2026 **Article DOI: 10.65150/EP-jsshrs/V2E10/2026-07**

ABSTRACT: Generative AI tools are increasingly embedded in higher education, where English majors are frequent users. Although they reduce cognitive load and boost efficiency, over-reliance may weaken independent thinking and logical reasoning, both core to critical thinking. This study examines how AI use influences critical thinking among undergraduate English majors at a university in Shanghai, focusing on usage frequency and question types. A mixed-methods design combines a questionnaire with semi-structured interviews. Questionnaire data are analyzed in SPSS using descriptive statistics, correlation, and regression; interviews capture perceived cognitive effects. The findings identify how frequency and question types relate to different dimensions of critical thinking, revealing both facilitating and constraining effects. Theoretically, the study extends research on AI in humanities education; practically, it informs strategic AI integration and critical thinking cultivation in English language teaching and digital transformation.

KEYWORDS: generative AI, AI tool use, critical thinking, English majors, mixed-methods research.

1 INTRODUCTION

Generative AI tools are rapidly reshaping higher education, transforming pedagogy, assessment, and institutional arrangements across universities (Altamirano et al., 2026; McDonald et al., 2025). Within the higher education sector, university students have become the most active participants and practitioners of this technological transformation, as they demonstrate the highest level of acceptance towards new technologies. Undergraduate students majoring in English are a prime example of this; they frequently utilize AI tools to carry out learning activities such as translation assistance and text polishing, which reduces mechanical workload and improves efficiency (Qi et al., 2024). However, this convenience is accompanied by potential risks and challenges. If students become overly reliant on AI-generated answers, delegating cognitive processes that should be undertaken independently to machines, this may weaken their capacity for independent reasoning, in-depth analysis and rational judgement, thereby hindering the development of higher-order thinking skills. Yet it is precisely these abilities that constitute the core competency required of English language professionals: critical thinking (Cai & Chen, 2025). Critical thinking involves the logical analysis of complex texts and the ability to construct innovative arguments in academic writing; it forms a vital foundation of the core competitiveness of English language and literature students (Wen et al., 2009; Facione, 1990). Therefore, examining the impact of AI tools on the critical thinking of undergraduate students in English language and literature holds significant practical importance.

This study makes both theoretical and practical contributions. Theoretically, it extends empirical research on AI and critical thinking by focusing on English-major undergraduates in discipline-specific contexts such as AI-assisted translation and text polishing, rather than treating university students as a homogeneous group. By introducing usage frequency and question types, it constructs a more refined analytical framework that moves beyond single-dimensional measurement. Its mixed-methods design, combining questionnaire surveys with semi-structured interviews, also provides both large-scale evidence and contextual depth, offering methodological insights for future research. Practically, the findings can help English language teachers and administrators offer targeted guidance on appropriate AI use and mitigate cognitive outsourcing. They can also help students adopt a rational stance toward AI and adjust their usage strategies according to their own frequency and question types, thereby sustaining and developing their critical thinking.

License: This is an open access article under the CC BY 4.0 license:

https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Yanhui Chen

2 LITERATURE REVIEW

2.1 AI Tools in English Language Teaching

With the rapid development of AI technology, various AI tools have gradually entered the field of education, demonstrating particularly broad prospects for application in English language teaching. Currently, AI tools commonly used in English teaching fall into three categories: generative conversational bots such as ChatGPT and DeepSeek for essay polishing and idea generation; machine translation tools such as DeepL and Google Translate; and intelligent writing assistants such as Grammarly for grammar detection and improvement. Existing research generally agrees that these tools enhance the efficiency of English language learning. For instance, Kohnke et al. (2023) note that ChatGPT provides immediate feedback and personalized support, particularly in writing and speaking practice. Xu (2024) finds that AI-enhanced writing instruction can broaden students' thinking depth. Sun et al. (2026) report that students using generative AI tools demonstrate significantly superior critical thinking and self-directed learning abilities than non-users, partly because AI reduces the burden of low-cognitive-load tasks.

However, challenges persist. Over-reliance is a primary concern: students accustomed to directly generating answers may experience declines in independent thinking and logical argumentation (Cai & Chen, 2025). Recent research demonstrates that generative AI plays a dual role in misinformation-related tasks, with inconsistent mitigation outcomes and a need for standardized assessment and regulatory measures (Park & Nan, 2026). A recent systematic review identifies dual impacts of generative AI and LLMs: they may boost EFL learners' critical thinking yet risk undermining it via over-reliance and technical-pedagogical flaws (Liu et al., 2025). Overall, AI tools present a "double-edged sword" effect, and research targeting English majors specifically remains limited.

2.2 Critical Thinking: Definition and Influencing Factors

Critical thinking is broadly understood through two dimensions. From a cognitive-skills perspective, it involves discerning the essence of objects and propositions and articulating independent, logical views (Qi et al., 2024). From a dispositional perspective, it is a form of rational reflection to determine what to believe or do (Ennis, 2015). Cai and Chen (2025) further define it as a composite ability encompassing both critical spirit and critical skills. Synthesizing these views, this study defines critical thinking as the cognitive ability and mental disposition to actively engage in logical analysis, rational evaluation, and reflective judgment when confronted with information or problems.

Three categories of influencing factors emerge from the literature. First, individual internal factors, such as gender, academic discipline, academic performance, and cognitive style, shape baseline critical thinking levels. Second, educational environmental factors matter: when teachers present content as neutral or fail to expose students to diverse perspectives, independent critical consciousness is difficult to develop (Sun, 2026). Third, and most relevant to this study, technology-use factors, particularly usage frequency and question types, exert significant influence, as detailed below.

2.3 Relationship Between AI Tools and Critical Thinking

Existing research has not reached consensus on how AI-tool use relates to critical thinking, with viewpoints falling into three broad categories: positive effects, negative effects, and complex relationships.

Pro-AI researchers hold that AI reduces cognitive burden for low-level tasks and frees cognitive resources for higher-order thinking. Empirical evidence shows generative-AI users exhibit better critical-thinking and self-directed learning, as AI handles repetitive work including vocabulary searching and literature sorting (Qi et al., 2024). AI chatbots deliver instant, personalized feedback to boost learning efficiency, and AI-supported writing pedagogy can advance students' emotional authenticity, critical-creative thinking, and linguistic appropriateness to strengthen argument construction (Kohnke et al., 2023; Xu, 2024). From this view, AI extends human thinking boundaries.

By contrast, scholars highlighting negative impacts warn against over-reliance. Generative AI's automatic content generation may rob learners of meaningful information-processing experiences. Without solid knowledge and discernment, students are vulnerable to AI-produced misinformation (Liu & Dong, 2025). Besides, AI performs unevenly in evaluating complex logical reasoning and argument depth across thematic tasks, implying inherent limitations for sophisticated cognitive work (Leng et al., 2024). Blind acceptance of AI outputs may erode independent-thinking capacities.

The complex-relationship perspective integrates the above stances. Based on over 10 000 college-student questionnaires, Cai and Chen (2025) proposed a U-shaped-effect hypothesis: low-frequency AI use inhibits critical thinking, yet high-frequency sustained use alleviates such adverse influences, suggesting usage manner outweighs mere frequency. Furthermore, Qi et al. (2024) found that low-cognitive-level closed-ended questions showed significant positive associations with critical-thinking, while open-ended questions, which students rarely employed, exhibited only weak correlations with such growth. These findings imply that task type and questioning strategies determine whether AI serves as a booster or a crutch for learners.

Despite notable progress in existing research on AI tools and critical thinking, two key research gaps can be identified. Firstly, current studies mostly target general undergraduates rather than English majors, whose language- and culture-oriented learning imposes unique critical-thinking demands. Moreover, AI-usage indicators are often oversimplified, with insufficient differentiation between usage frequency and question types. Accordingly, this study focuses on English majors to investigate AI's influence on critical thinking across the two dimensions of usage frequency and question types to address the gaps in existing research.

3 METHODOLOGY

3.1 Research Design

The data for this study were primarily collected through a questionnaire survey, supplemented by standardized scales to measure core variables, thereby ensuring the authenticity, validity and relevance of the data. The specific data sources and related notes are as follows:

License: This is an open access article under the CC BY 4.0 license:
https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen

3.1.1 Research Participants

This study adopted a mixed sampling strategy integrating convenience sampling and stratified sampling to recruit English-major undergraduates from a Shanghai-based university. Stratification by academic year and gender was implemented to achieve balanced sample distribution across grade levels, satisfying statistical assumptions for independent-variable analysis and enhancing sample representativeness. Among 66 collected questionnaires, 54 valid copies were retained after eliminating invalid returns marked by incomplete information, repetitive replies and abnormally short completion duration, corresponding to an 81.8 % valid response rate. The basic information of the sample is as follows: the distribution by year of study was 6 first-year students (11.1%), 12 second-year students (22.2%), 15 third-year students (27.8%) and 21 fourth-year students (38.9%); the gender distribution comprised 8 males (15%) and 46 females (85%), which is broadly consistent with the current gender ratio among undergraduate students majoring in English in China, indicating that the sample is reasonably representative.

3.1.2 Questionnaire Design

The first section covers demographic information and academic background, including variables such as gender, year group, and English proficiency. Part Two examines the use of AI tools, primarily measuring the frequency of use, types of tools commonly used, usage scenarios, and question types (referencing the classification framework of four types of cognitive questions proposed by Qi et al., 2024). Part Three comprises a critical thinking scale, adapted from the Chinese version of the California Critical Thinking Disposition Inventory (CCTDI) and appropriately revised to suit the context of English language studies, comprising a total of 15 items. The scale employs a five-point Likert scale. In this study, the scale's internal consistency reliability, as measured by Cronbach's α, was 0.916, indicating good reliability.

3.1.3 Interview Data Collection

To supplement quantitative questionnaire data, four participants were selected via typical sampling from 54 questionnaire respondents for semi-structured interviews. Selection considered AI-use frequency, critical-thinking scores, and year groups to guarantee sample heterogeneity and representativeness. Face-to-face interviews (10-20 min each) were fully recorded and transcribed with participant consent, generating 4026-word textual data. Interviews covered AI-use experiences, perceived critical-thinking shifts, AI-related challenges and coping strategies, and AI-use suggestions for English teaching.

3.2 Measurement of Variables

Centering on the research question concerning AI tools' effects on English-literature undergraduates' critical thinking, this study specifies independent, dependent and control variables with clear conceptual definitions and operational measurements to standardize research procedures and secure result reliability.

Independent variables include AI-tool use frequency and question-type characteristics. AI-use frequency measures how often students deploy AI for English-learning tasks, captured by a six-point ordinal scale in the questionnaire, with higher scores representing more frequent usage. Drawing on Qi et al. (2024), question types are adapted to English-learning contexts and coded into four categories by cognitive load and openness: low-cognitive closed-ended, low-cognitive open-ended, high-cognitive closed-ended, and high-cognitive open-ended. Each category's usage frequency is measured via five-point Likert items; the averaged score of the two high-cognitive categories serves as the indicator for deep questioning behavior.

The dependent variable, critical thinking, denotes students' cognitive capacities and dispositions for logical analysis, rational evaluation and reflective judgment when processing English-learning-related information. This study adopts a context-adapted 15-item Chinese version of the California Critical Thinking Disposition Inventory (CCTDI). Anchored on a five-point Likert scale, total scores ranging 15-75 reflect overall critical-thinking competence.

Gender, year group, and academic proficiency are set as control variables to mitigate confounding effects. Gender is coded as a dichotomous variable. Academic year is treated as an ordinal variable corresponding to students' undergraduate stages. Academic proficiency is operationalized by students' highest English-language certification, with ordinal codes increasing alongside examination difficulty.

3.3 Data Analysis

Questionnaire data are processed with SPSS 26.0. Descriptive statistics generate frequencies, percentages, means and standard deviations to illustrate sample profiles, AI-use status and critical-thinking score distributions. Correlation analysis preliminarily explores relationships between AI-related variables and critical-thinking performance. One-way ANOVA together with LSD post-hoc tests and Cohen's d effect-size calculation examine critical-thinking disparities across AI-use-frequency subgroups. An independent-samples t-test further compares critical-thinking scores grouped by the median of high-cognitive questioning frequency. The transcribed interview data undergo thematic analysis to extract core themes for triangulating quantitative findings.

4 RESULTS AND DISCUSSION

4.1 AI Tool Usage among English Major Undergraduates

To gain a comprehensive understanding of the general situation regarding the use of AI tools by undergraduate students majoring in English Language and Literature, this study conducted a descriptive statistical analysis of the valid sample. The analysis focused primarily on two key dimensions: the frequency of AI tool usage and the types of queries. It also included comparative analyses across groups defined by students' academic performance.

4.1.1 Distribution of Basic Information

As shown in Table 1, the proportion of female students in the sample (85.2%) is significantly higher than that of male students (14.8%), which is consistent with the current gender ratio among undergraduate students majoring in English in China. In terms of year group distribution,

License: This is an open access article under the CC BY 4.0 license:
https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen

fourth-year students constitute the largest group (38.9%), whilst first-year students are the smallest (11.1%); overall, there is a trend of increasing numbers as the year group progresses. In terms of academic attainment, the majority of students had passed the College English Test Band 6 (CET-6) (38.9%), followed by those who had passed the College English Test Band 4 (CET-4) and the Professional English Test Band 4 (27.8% each); fewer students had passed the Professional English Test Band 8 (3.7%), which is consistent with the progressive development of language proficiency typically observed at the undergraduate level in English language programmers.

Table 1: Demographic characteristics of participants (N=54)

Items	Categories	Number	Percent
Year	First year Second year	6	11.1%
		12	22.2%
	Third year	15	27.8%
Items	**Categories**	**Number**	**Percent**
	Fourth year	21	38.9%
Gender	Male	8	14.8%
	Female	46	85.2%
Academic performance	Failed all subjects	1	1.9%
	CET-4	15	27.8%
	CET-6	21	38.9%
	TEM-4	15	27.8%
	TEM-8	2	3.7%

4.1.2 Frequency distribution of AI tool usage

Regarding the frequency of AI tool usage among English major undergraduates, the survey results show a relatively concentrated distribution pattern. Among the 54 valid respondents, the frequency distribution is presented in Table 2.

Table 2: Frequency distribution of AI tool usage among English majors

Frequency of AI use	Number of people selected	Percent
Several times a day	8	14.8%
About once a day	14	25.9%
3–5 times a week	16	29.6%
1–2 times a week	11	20.4%
2–3 times a month	4	7.4%
Once a month or less	1	1.9%

As indicated in Table 2, the overall frequency of AI tool usage among undergraduate English majors is relatively high. Among them, users who use such tools 3–5 times a week accounted for the highest proportion (29.6%), followed by those who use them approximately once a day (25.9%); together, these two groups accounted for over half of the total. Students who use them multiple times a day accounted for 14.8%, whilst those who use them 1–2 times a week accounted for 20.4%. In contrast, the proportion of infrequent users, those using such tools 2–3 times a month or less, was extremely low, totaling just 9.3%. Overall, 90.7% of students maintain the habit of using AI tools at least once a week. This indicates that AI tools have become deeply integrated into the daily studies of undergraduate English majors, serving as indispensable aids in learning scenarios such as vocabulary acquisition, text translation and writing assistance. High-frequency, routine use has become the predominant characteristic of usage within this group.

This finding is generally consistent with Qi et al. (2024), who reported that over 70% of university students use AI tools weekly. However, the proportion of daily users in this study (40.7%) is higher than that in Qi et al.'s sample, possibly due to the nature of English language tasks (e.g., frequent translation and proofreading) that invite regular AI use.

4.1.3 Distribution of question types

This study categorizes student questions when using AI tools into four categories: low-cognitive closed-ended, low-cognitive open-ended, high-cognitive closed-ended, and high-cognitive open-ended. A five-point Likert scale (1 = never, 5 = always) measured the frequency of each question type. The results are shown in Table 3.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen

As seen in Table 3, the average score for low-cognitive-closure questions was the highest (4.28 points), indicating that students most frequently used AI for basic tasks such as word lookup and grammar checking; The average score for high-cognitive, open-ended questions was the lowest (2.12 points), suggesting that students make less use of AI for in-depth critical thinking, debate or argument evaluation; scores for low-cognitive, open-ended and high-cognitive, closed-ended questions fell within the middle range. This distribution pattern indicates that students currently view AI more as a 'provider of answers' rather than a 'partner in critical thinking', and their usage patterns have yet to fully realize AI's potential value in expanding the boundaries of their thinking.

Table 3: Mean scores of different question types in AI tool use

Question type	Mean	SD	Description
Low-cognitive closed-ended	4.28	0.85	Word lookup, translation, grammar checking, etc.
Low-cognitive open-ended	3.19	1.08	Recommended learning resources, etc.
High-cognitive closed-ended	2.91	0.96	Analysis of grammatical phenomena, proofreading, etc.
High-cognitive open-ended	2.12	1.15	Evaluation of arguments, debating, etc.

This pattern mirrors Qi et al.'s (2024) finding that closed-ended questions dominate student-AI interactions. Notably, the mean score for high-cognitive open-ended questions in this English major sample (M=2.12) is even lower than the multi-disciplinary average reported by Qi et al. (M=2.45), suggesting that the "standard answer" tradition in humanities disciplines may further discourage students from posing exploratory, open-ended queries.

4.2 Critical Thinking Skills of English Major Undergraduates

4.2.1 Overall Level of Critical Thinking

As presented in Table 4, the mean total score for critical thinking among English major undergraduates was 3.29, with a standard deviation of 0.31, a median of 3.27 and a mode of 3.33. The total scores ranged from 2.53 to 3.73, indicating a certain degree of variation between individual students. Based on the total score for the 15-item test (scoring range 15–75), the mean total score for the sample was 49.35 (SD = 4.65), placing it at an intermediate level. The overall distribution was relatively concentrated (skewness coefficient –0.21), indicating that most students scored slightly above the median, but there was no obvious ceiling effect.

Table 4: Descriptive statistics for the overall score in critical thinking (N=54)

Indicators	Mean	SD	Median	Mode	Minimum	Maximum	Skewness
Total score（15-75）	49.35	4.65	49	48	38	56	-0.21
average score（1-5）	3.29	0.31	3.27	3.2	2.53	3.73	-0.21

4.4.2 Item-level Score Distribution

To gain a better understanding of students' performance across the various dimensions of critical thinking, the scores for each of the 15 items were analyzed individually; the results are shown in Table 5.

Table 5: Descriptive statistics of scores for each item of critical thinking (N=54)

Items	Question content	mean	SD	Percentage
10	When faced with a new idea, I first consider whether there is sufficient evidence to support it	3.78	0.72	68.50%
11	I am able to identify logical flaws in articles or arguments	3.52	0.68	55.60%
12	I am in the habit of analyzing the same issue from multiple perspectives	3.61	0.7	61.10%
13	Even when I agree with a viewpoint, I actively seek out counter-evidence.	3.2	0.85	37.00%
14	Before reaching a conclusion, I carefully scrutinize my reasoning process.	3.33	0.78	44.40%
15	I can clearly explain to others the reasons behind my conclusions.	3.41	0.74	48.10%
16	When it comes to content generated by AI tools, I actively verify its accuracy.	3.15	0.92	35.20%

License: This is an open access article under the CC BY 4.0 license:
https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen

17	When others express views that differ from my own, I listen carefully and consider their validity.	3.67	0.69	64.80%
18	I enjoy challenging myself by thinking about complex problems that have no standard answers	2.91	0.85	22.20%
19	When reading English texts, I pay attention to whether the author has used biased language	3.24	0.8	40.70%
20	I reflect on whether I have any blind spots or biases in my thinking	3.3	0.76	42.60%
21	I am willing to raise questions during discussions, even if my views may not be accepted by others	3.08	0.88	31.50%
22	I believe that making judgements based solely on intuition is often unreliable	3.45	0.73	51.90%
23	Before making a decision, I gather as much information as possible	3.65	0.68	63.00%
24	I regularly reflect on whether my use of AI tools is appropriate	2.84	0.91	20.40%

The three items with the highest scores were, in order: 'When faced with a new viewpoint, I first consider whether there is sufficient evidence to support it' (3.78), 'When others express views that differ from my own, I listen carefully and consider their validity' (3.67), and 'Before making a decision, I gather as much information as possible' (3.65). This indicates that undergraduate students majoring in English perform well in terms of verifying information, listening to others and making prudent decisions, demonstrating foundational critical thinking skills. Furthermore, the items "I am accustomed to analyzing problems from multiple angles" (3.61 points) and "I can identify logical flaws" (3.52 points) also received high scores, suggesting that students have begun to develop an awareness of multi-perspective thinking and logical analysis in their English studies.

(1) Weak Areas: Proactive Challenge and Self-Reflection

The three items with the lowest scores were, in order: "I regularly reflect on whether my use of AI tools is appropriate" (2.84 points), "I enjoy thinking about complex problems with no standard answers" (2.91 points) and "I am willing to raise questions during discussions" (3.08 points). This suggests that students generally lack reflection on their own technology usage, are reluctant to actively engage with open-ended or uncertain problems, and are hesitant to express dissenting views in discussions.

(2) Moderate Awareness of AI Content Verification

The score for "I actively verify the accuracy of content generated by AI tools" was 3.15, placing it at an intermediate level. While most students do not place complete trust in AI outputs, the habit of actively verifying such content has not yet become widespread. Given that AI sometimes produces content that appears correct but is in fact erroneous (the so-called 'hallucination' phenomenon), this result suggests that students are at risk of being misled.

Overall, students perform reasonably well in terms of foundational critical thinking, but there is still considerable room for improvement in terms of initiative in higher-order thinking, reflection on technology use, and the courage to question.

4.3 Impact of AI Tools on the Critical Thinking of English Majors

This section systematically examines the specific impact of AI tool usage on the critical thinking of undergraduate English majors, based on questionnaire survey data and semi-structured interview materials, from the dual perspectives of usage frequency and question types. Usage frequency reflects the 'quantity' of technology use, whilst question types reflect its 'quality'. The analysis employs correlation analysis, analysis of variance (ANOVA) and independent samples t-tests, supplemented by interview materials for triangulation.

4.3.1 Analysis of the Impact of Usage Frequency

First, a correlation analysis was conducted between the frequency of AI tool use (a 6-point ordinal variable, coded 1–6) and the mean critical thinking score. The results indicate a weak positive correlation between frequency of use and critical thinking scores ($r = 0.09$, $p = 0.52$), which did not reach statistical significance. This suggests that simply increasing the frequency of AI tool use does not necessarily lead to an increase or decrease in critical thinking; there is no simple linear relationship between the two.

To further examine differences between frequency groups, the usage frequencies were categorized into three groups: the low-frequency group (2–3 times per month or less, including codes 1–2, n = 5), the medium-frequency group (1–2 times per week to 3–5 times per week, codes 3–4, n = 27), and the high-frequency group (approximately once a day or more, codes 5–6, n = 22). A one-way analysis of variance (ANOVA) was conducted with the mean critical thinking score as the dependent variable; the results are shown in Table 6.

Table 6: Comparison of critical thinking scores across different usage frequency groups

Frequency of use group	Sample size (n)	Mean	SD	F	p
Low-frequency group	5	3.21	0.28	0.48	0.62

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen

Medium-frequency group	27	3.26	0.3
High-frequency group	22	3.34	0.33

As shown in Table 6, the mean critical thinking scores for all three groups increased gradually with higher AI tool use frequency (3.21, 3.26, and 3.34, respectively). However, the differences between groups were not statistically significant (F (2, 51) = 0.48, p = 0.62). This suggests that, within the study sample, the frequency of AI tool use does not lead to significant differences in critical thinking skills.

This finding does not fully align with the 'U-shaped relationship' hypothesis proposed by Cai and Chen (2025) based on 10,964 university student questionnaires. There are three possible reasons for this: Firstly, the sample size of low-frequency users in this study was too small (only five individuals, accounting for 9.3%), making it difficult to effectively capture the potential negative effects of 'sporadic reliance'. Secondly, usage frequency itself is not the core mechanism influencing thinking; the key lies in the level of cognitive engagement and the strategies employed during use. Thirdly, the sample in this study consisted entirely of undergraduate students majoring in English, whose academic tasks inherently require a certain level of linguistic analysis and critical thinking, which may have mitigated the impact of mere changes in frequency to some extent.

Interview data provide concrete evidence for the above explanations. One high-frequency user (TEM-4, fourth-year undergraduate, using it multiple times daily) stated:

"I use AI every day to edit essays and search for academic literature, but I think things through myself first, then compare my thoughts with the AI's suggestions to see where it is right and where it is wrong. Sometimes the AI's explanations aren't accurate, and I can spot its flaws."

In contrast, another medium-frequency user (CET-6, second-year undergraduate, 3–5 times a week) admitted:

"Sometimes I'm too lazy to think and just ask the AI; its answers seem quite reasonable, so I don't question them much. However, occasionally I feel this isn't quite right, I feel I'm becoming lazy."

These contrasting cases echo Liu and Dong's (2025) warning about "cognitive outsourcing": the first student maintains intellectual autonomy while the second exhibits early signs of cognitive laziness, illustrating that frequency alone does not determine outcomes, what matters is whether the user engages metacognitively with AI outputs.

Although the two students had similar usage frequencies (both high or medium-high), their self-assessments of critical thinking and actual performance differed markedly due to variations in cognitive engagement and reflective habits. This suggests that the impact of usage frequency on critical thinking is indirect, moderated by individual cognitive engagement, reflective habits and usage strategies. Consequently, in teaching interventions, simply encouraging students to 'use more' or 'use less' AI tools may have limited effectiveness; greater attention should be paid to 'how to use' them.

4.3.2 Analysis of the Impact of Question Types

The type of question asked reflects the depth of cognitive engagement demonstrated by students when using AI tools. This study categorized question types into four groups: low-cognitive closed-ended, low-cognitive open-ended, high-cognitive closed-ended and high-cognitive open-ended. The average frequency of use (on a scale of 1–5) for each student across these question types was calculated, and a correlation analysis was conducted with the mean critical thinking score; the results are shown in Table 7.

(1) Correlation Analysis

Table 7: Pearson correlation matrix between question types and critical thinking scores (N=54)

Variables	**1**	**2**	**3**	**4**	**5**
1. Critical thinking	1				
2. Low cognitive closure	-0.28*	1			
3. Low cognitive openness	0.11	0.34**	1		
4. High cognitive closure	0.42**	-0.05	0.29*	1	
5. High cognitive openness	0.53**	-0.18	0.21	0.48**	1

***Note.* *p* < .05, **p* < .01.**

Table 7 reveals a moderate positive correlation between high-cognitive open-ended questions (e.g., argument evaluation, requesting dual-sided viewpoints, debate, and perspective exploration) and critical thinking (r=0.53, p<0.01). Students posing such reflective AI questions more often tend to have stronger critical thinking. Secondly, high-cognitive, closed-ended questions (such as analyzing grammatical phenomena, requesting revisions and justifying them, or extracting key information from texts) also show a significant positive correlation with critical thinking (r=0.42, p<0.01), but the effect is slightly weaker than that of the high cognitive openness type. Furthermore, closed-ended questions requiring little critical thinking (such as word look-ups, translations and simple grammar checks) are significantly negatively correlated with critical thinking (r=-0.28, p<0.05). This suggests that over-reliance on AI to perform basic tasks with fixed answers may hinder the development of critical thinking, or even have a stifling effect. Finally, low-cognitive, open-ended questions (such as those asking for recommended learning resources or examples of everyday expressions) are not significantly associated with critical thinking (r=0.11, p>0.05). These correlational patterns align with Qi et al.

License: This is an open access article under the CC BY 4.0 license:
https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen

(2024), who also found that closed-ended questions were negatively associated with critical thinking (r=-0.31 in their study). However, while Qi et al. reported a positive correlation between low-cognitive open-ended questions and critical thinking (r=0.35), that relationship was not significant in this English major sample (r=0.11), possibly due to disciplinary differences in how "open-ended" tasks are conceptualized.

(2) Comparison of groups based on the frequency of high-cognitive, open-ended questions

To further validate the aforementioned correlation, students were divided into high- and low-frequency groups based on the median frequency of high cognitive openness questions (2.0 points). The mean critical thinking score for the high-frequency group (N=26) was 3.52 (SD=0.27), whilst that for the low-frequency group (N=28) was 3.08 (SD=0.24). The results of the independent samples t-test showed a significant difference between the two groups ($p < 0.05$), with an effect size of Cohen's D = 1.73, indicating a very large effect. This result strongly suggests that the frequency with which students ask high-cognitive-openness questions when using AI tools is a key indicator for distinguishing between high and low levels of critical thinking.

(3) Evidence from interviews

The interviews further revealed the micro-mechanisms by which question types influence critical thinking. A student who frequently uses high-cognitive-openness questions (TEM-4, third-year) shared her strategy:

"I get ChatGPT to play the role of the opponent and debate a literary viewpoint with me. For example, if I put forward an interpretation of a novel's theme, it will deliberately find counterexamples to challenge me. I have to reorganise my evidence and patch up logical gaps; this process is far more effective than just thinking it through on my own. Sometimes I also ask it to guide me step by step using Socratic questioning; it really helps me uncover issues I hadn't noticed before."

Another student with a similarly high score (TEM-4, fourth-year) added:

"When writing essays, I don't ask the AI to write paragraphs for me directly. Instead, I first outline my structure and arguments, then ask it, 'What are the flaws in this argument?' or 'Is there any counter-evidence I've overlooked?' In this way, the AI acts as a sharpening stone for my thinking."

In contrast, a student who primarily uses low-cognitive, closed-ended questions (CET-4, second-year) stated:

"I generally just use the AI to look up words, translate sentences and check for grammatical errors; I rarely ask it questions that require thought. Sometimes I want to ask something more profound, but I don't know how to phrase it, or I'm afraid I won't understand the answer."

Another student who rarely uses high-cognitive open-ended questions (CET-6, third-year) admitted:

"I see AI as just a tool; as long as it gives me the correct answer, that's fine. Debating with it? It feels a bit odd, it's not a real person, so why should I argue with a machine?"

The above interview excerpts indicate that students who view AI as a 'discursive partner' rather than an 'answer machine' are able to proactively pose high-cognitive, open-ended questions, thereby honing their logical reasoning, evidence evaluation and argument reconstruction skills through interaction with AI; whereas students who treat AI merely as an 'answer provider' are prone to falling into a state of passive acceptance and intellectual inertia.

(4) Summary

Based on a synthesis of relevant analyses, comparative group studies and interview evidence, the following conclusions can be drawn: Firstly, the type of question asked is a better predictor of critical thinking levels than the frequency of use. The frequency of use is not the key factor; what matters is 'what is asked' and 'how one interacts with AI'. In addition, high-cognitive, open-ended questions are an effective pathway to fostering critical thinking. Such questions compel students to evaluate, compare, argue and reflect beyond mere information retrieval, thereby activating higher-order cognitive processes. Moreover, the overuse of low-cognitive, closed-ended questions may pose cognitive risks. When students primarily use AI to perform basic cognitive tasks on their behalf, the cognitive outsourcing effect may undermine their ability to think independently. More importantly, improving the quality of students' questions is a viable direction for pedagogical intervention. English language teachers can guide students to shift from 'using AI to find answers' to 'using AI to practice thinking' by demonstrating higher-order questioning and designing critical thinking tasks. Taken together, these findings support Liu and Dong's (2025) assertion that AI is a "double-edged sword" for critical thinking: its effects depend not on whether students use it, but on how they engage with it, particularly whether they ask high-cognitive open-ended questions that stimulate reflective judgment.

4.4 Implications of the Study

4.4.1 Implications for English Majors

This study identifies a significant positive correlation between high-cognitive open-ended questioning and critical thinking (r=0.53). Nevertheless, participants' mean score for this question type stood at only 2.12, much lower than the 4.28 recorded for low-cognitive closed-ended questions. Students also scored poorly on self-monitoring of appropriate AI use (M=2.84). This suggests that students need to adjust their usage behavior in two respects: First, learners should upgrade query structures: after completing routine tasks such as vocabulary checks or essay polishing, they may raise evaluative follow-up questions to turn instrumental AI use into dialogic critical practice. Second, students need metacognitive monitoring, regularly reflecting on whether AI has replaced their independent reasoning to avoid thought substitution.

4.4.2 Implications for English Language Teachers

Many students rarely formulate high-order questions because they lack relevant strategies and worry about comprehending AI responses. Instead of simply permitting or banning AI use, teachers should embed AI questioning training into instruction. They can demonstrate advanced questioning in reading, writing and literary analysis classes to help students link question design to thinking depth. In addition, human-AI collaborative tasks, such as AI-supported argument evaluation, can be assigned. Students form original viewpoints, engage AI as an opposing debater, and submit dialogue records plus reflections, enabling teachers to diagnose thinking weaknesses and offer targeted feedback.

License: This is an open access article under the CC BY 4.0 license:
https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen

4.4.3 Implications for Curriculum Design

Low student scores on embracing open-ended complex thinking (M=2.91) and voicing objections in discussions (M=3.08) suggest that overemphasis on standard answers in assessment may discourage critical inquiry. Curriculum evaluation could reduce objective items and expand process-oriented assessments including essays, projects and oral debates to reward critical exploration. Universities should also integrate AI literacy into academic integrity training, establishing context-specific AI usage guidelines. Such policies reject blanket prohibition while limiting over-reliance, cultivating a learning culture grounded in critical and rational technology use.

5 CONCLUSION

This study identifies a non-linear association between AI tool use and critical thinking among English majors. Although 90.7% of participants use AI weekly or more, usage frequency bears no significant linear correlation with critical thinking ($r=0.09$, $p=0.52$; $F=0.48$, $p=0.62$). Interview evidence reveals that cognitive engagement and reflective habits moderate its effects, refining the U-shaped hypothesis for this population and highlighting that use quality matters more than use quantity. Question type serves as a stronger predictor. High-cognitive open-ended questions show a moderate positive correlation with critical thinking ($r=0.53$, $p<0.01$), and the high-frequency group of such questions achieves significantly higher critical thinking scores (Cohen's $d=1.73$). High-cognitive closed-ended questions also correlate positively ($r=0.42$, $p<0.01$). Conversely, low-cognitive closed-ended questions, the most commonly adopted type ($M=4.28$), correlate negatively with critical thinking ($r=-0.28$, $p<0.05$), bringing risks of cognitive outsourcing. The sample demonstrates a moderate overall critical thinking level ($M=3.29$). Participants perform well in evidence evaluation and accepting diverse views, yet show weaknesses in tackling open complex tasks, metacognitive monitoring of AI use, and voicing doubts in discussions.

This research has several constraints. The sample, consisting of 54 questionnaires and four interviewees from one university, limits generalizability. Critical thinking is measured solely via a self-reported adapted CCTDI scale, vulnerable to social desirability bias. The small, short semi-structured interviews lack teachers' perspectives. Besides, question types rely on self-reported frequency rather than real AI dialogue texts, which may introduce subjective bias.

Future work may recruit participants across diverse universities and balance year-group distribution. Multiple measurements including performance-based tasks can complement self-report scales. Researchers may expand interview samples, incorporate teacher interviews and classroom observation. Real student-AI conversation transcripts can also be collected for objective coding of question cognitive levels to validate the link between questioning quality and critical thinking.

REFERENCES

1) Altamirano, S. F. S., Maquen-Niño, G. L. E., Roldán, C. M. G., Herrera, A. P., & Cárdenas, D. M. C. (2026). Generative artificial intelligence in higher education: A systematic review of perceptions, implementation and pedagogical transformation. *Review of Education*, 14, e70152. https://doi.org/10.1002/rev3.70152
2) Cai, Y. W., & Chen, Z. (2025). The double-edged sword: The impact of generative artificial intelligence on college students' critical thinking. *Journal of Huazhong Normal University (Humanities and Social Sciences)*, **64**(6), 146-159. https://doi.org/10.19992/j.cnki.1000-2456.2025.06.015
3) Ennis, R. H. (2015). Critical thinking: A streamlined conception. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 31-47). Palgrave Macmillan. https://doi.org/10.1057/9781137378057_2
4) Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction* (The Delphi Report). California Academic Press.
5) Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 1–5.
6) Leng, J., Lu, H. H., & Dai, L. (2024). Empowering critical-thinking assessment with generative artificial intelligence: An applied experiment based on ChatGPT. *Modern Distance Education Research*, **36**(6), 102-111.
7) Liu, F., & Dong, Y. (2025). Why is critical thinking increasingly important in the age of generative AI? *Media in Foreign Language Instruction*, **47**(1), 3-10+113.
8) 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
9) McDonald, N., Johri, A., Ali, A., & Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines.*Computers in Human Behavior: Artificial Humans*, *3*, 100121. https://doi.org/10.1016/j.chbah.2025.100121
10) Park, S., & Nan, X. (2026). Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation. *AI & SOCIETY*, *41*, 1501-1515. https://doi.org/10.1007/s00146-025-02620-3
11) Qi, J., Xu, Y. R., Liu, J. A., & Xue, K. (2024). The impact of generative artificial intelligence tools on college students' critical thinking and autonomous learning ability. *e-Education Research*, **45**(12), 67-74. https://doi.org/10.13811/j.cnki.eer.2024.12.009
12) Sun, X., Liu, C., Li, P., & Cai, Z. (2026). Effects of Generative AI Tools On Students' Critical Thinking: A Three-Level Meta-Analysis. *Educational Psychology Review*, **38**(84). https://doi.org/10.1007/s10648-026-10180-1

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Yanhui Chen

13) Wen, Q. F., Wang, J. Q., Zhao, C. R., Liu, Y. P., & Wang, H. M. (2009). Developing a conceptual framework for assessing Chinese university students' critical thinking skills. *Foreign Language World*, **30**(1), 37-43.
14) Xu, L. (2024). AI-enhanced English writing instruction and students' thinking development. *Foreign Language World*, (3), 58–65.

License: This is an open access article under the CC BY 4.0 license:
https://creativecommons.org/licenses/by/4.0/
Corresponding Author: Yanhui Chen