

Exploring the Relationship Between Generative Artificial Intelligence Usage and the Development of University Students' Critical Thinking in Higher Education

Agustina^{1*}, Dwi Otik Kurniawati², Boy Lumoindong³, Irwan Suanto⁴

¹Politeknik Arutala Johana Hendarto Jakarta

²UIN Sunan Kalijaga Yogyakarta

³Universitas Klabat

⁴Universitas Islam Negeri Sultan Syarif Kasim Riau

*Corresponding Author: Agustina, agustmy@gmail.com

ARTICLE INFO

Keywords: Critical Thinking, Generative Artificial Intelligence, Students, Higher Education, Longitudinal Studies.

Received : 27, June

Revised : 29, July

Accepted: 25, August

©2026 Agustina, Kurniawati, Lumoindong, Suanto: This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

The increasingly widespread use of generative artificial intelligence (GenAI) in higher education has raised concerns about its influence on the development of students' critical thinking. This study aims to analyze the relationship between the intensity of GenAI use and changes in students' critical thinking skills. The study used a quantitative approach with a longitudinal correlational design on 186 students from three universities. Data was collected through GenAI usage questionnaires and critical thinking tests at the beginning and end of the semester. The analysis was performed using descriptive statistics, paired t-tests, Pearson correlation, and linear regression. The results showed an increase in critical thinking scores from 68.4 to 74.9, as well as a moderate positive relationship between the use of reflective GenAI and critical thinking development ($r=0.46$; $p<0.001$). The findings confirm the importance of the integration of directed, reflective, and evaluation-based GenAI in learning to improve the quality of higher education

INTRODUCTION

The development of generative artificial intelligence (GenAI), such as ChatGPT, Gemini, and Perplexity, has changed the way students obtain information, develop ideas, compose writing, and complete academic assignments. In higher education, GenAI can provide explanations quickly, generate alternative answers, provide feedback, and support personalized learning. However, this convenience also poses pedagogical problems because students can receive AI outputs without independently checking sources, evaluating arguments, or reflections. Van den Berg and du Plessis (2023) explain that GenAI can support learning planning and openness, but the resulting results should still be critically examined as they have the potential to contain bias and inaccuracies. On the other hand, experimental research by Essel et al. (2024) shows that the targeted use of ChatGPT in classroom activities can increase students' critical openness and reflective skepticism. This condition shows that GenAI does not automatically strengthen or weaken critical thinking; The impact depends on the purpose, usage pattern, and pedagogical assistance applied.

A similar phenomenon is found in the context of higher education in Indonesia. Akastangga et al. (2023) found that the use of ChatGPT has an impact on students' critical thinking skills, but reliance on instant answers has the potential to reduce the ability to evaluate information independently. Research by Refaldi et al. (2024) on 53 students of the Sepuluh Nopember Institute of Technology also shows that students view GenAI as a technology that helps academic productivity, while at the same time being aware of the risks to the originality of the work and the ability to think critically. Meanwhile, Jasmadi et al. (2024), through a study of 287 college students, found that AI can support data analysis, problem-solving, and decision-making, but also give rise to technology dependence and decrease active involvement in discussions. These findings confirm the urgency of this research because universities are not only responsible for improving students' ability to use technology, but also ensure that the use of technology still involves analysis, evaluation of evidence, drawing conclusions, and reflection as the main elements of critical thinking.

Theoretically, the relationship between the use of GenAI and critical thinking can be explained through two opposing mechanisms, namely GenAI as a cognitive scaffold and as a means of cognitive offloading. As a cognitive scaffold, GenAI can help students access diverse perspectives, compare alternative answers, obtain feedback, find weaknesses in arguments, and improve their thinking results. This kind of use encourages students to remain the main controller in the process of knowledge formation. In contrast, cognitive offloading occurs when students hand over cognitive activities, such as remembering, analyzing, assessing, and making decisions, to technology. Gerlich (2025) found a negative relationship between the frequency of AI use and critical thinking abilities mediated by increased cognitive offloading. In the Indonesian context, Lukman et al. (2023) also identified the risk of plagiarism, superficial understanding, and reduced critical thinking activities when students use AI answers as the final product. Thus, the intensity of AI use cannot be treated as

the only indicator; Reflective and evaluative orientation in its use needs to be considered.

Previous studies have shown inconsistent results. Putri and Panduwinata (2025) found a positive and significant effect of the use of AI on critical thinking skills in 279 students, showing that access to information, personalization of learning, and problem-solving support can strengthen analytical and evaluative skills. Fadholi (2025) also found that the use of ChatGPT is positively related to critical thinking and student decision-making when the convenience and usefulness of technology encourage productive use. These results are in line with Essel et al. (2024), who showed an increase in critical thinking skills in a group of students who used ChatGPT in a structured manner. However, Akastangga et al. (2023) asserted that dependence can decrease thinking independence, while Saraswati et al. (2023) identified changes in the tendency to laziness of thinking after the use of ChatGPT, although the effect was not statistically significant. These differences in results show that the relationship between GenAI and critical thinking is complex and influenced by the way students process, examine, and use AI outputs.

Although research on GenAI and critical thinking is starting to develop, there are a number of scientific gaps. Most Indonesian research still uses a cross-sectional design, conducted at a single university or study program, and relies on perception questionnaires given at one time. Akastangga et al. (2023), Refaldi et al. (2024), Putri and Panduwinata (2025), and Fadholi (2025), for example, have provided evidence on the relationship between the use of AI and critical thinking, but have not measured changes in students' abilities repeatedly in a learning period. Previous studies have also measured the use of AI more often in general and have not adequately distinguished reflective uses—such as comparing arguments, examining evidence, and revising answers—from passive uses of copying AI results. Even the experimental research of Essel et al. (2024) was carried out in a three-week intervention and recommended a study with a longer period. Therefore, longitudinal research is needed that combines measurements of GenAI use with critical thinking tests at the beginning and end of the semester to identify the development of abilities more precisely.

Based on these gaps, this study aims to analyze the relationship between the use of GenAI and the development of students' critical thinking skills in higher education. In particular, the research was directed to identify changes in students' critical thinking scores from the beginning to the end of the semester, measure the strength of the relationship between the use of GenAI and these changes, and analyze the contribution of reflective and evaluative use to the development of critical thinking. The study used a longitudinal correlational quantitative design involving 186 students from three universities, using a questionnaire using GenAI and a critical thinking test given at the beginning and end of the semester. Theoretically, this research is expected to clarify GenAI's position as a cognitive support tool as well as a source of cognitive offloading. Practically, the results of the research can be the basis for lecturers and universities to develop guidelines for the use of GenAI, design tasks that require

verification and reflection, and integrate AI without reducing students' independence of thinking.

LITERATURE REVIEW

Generative AI as a Cognitive Support

Generative Artificial Intelligence (GenAI) can serve as a cognitive scaffold when used to help students explore ideas, compare arguments, formulate questions, and evaluate information. Rizaldi et al. (2024) found that students who used ChatGPT obtained higher critical thinking scores than the control group in solving algebraic problems. Similar findings were shown by Ria and Yuda (2025), which showed the positive influence of the use of ChatGPT on critical thinking and computational thinking. In addition, Rahmawati et al. (2025) found that the use of ChatGPT contributes to the ability to structure questions based on clarity, depth, relevance, and originality. As such, GenAI has the potential to support high-level cognitive processes when used as an exploratory tool, rather than as a substitute for student thinking.

Reflective and Evaluative Use

GenAI's impact on critical thinking is mainly determined by the interaction patterns of its users. Nasr et al. (2025) distinguish AI-directed passive use from human-AI collaboration that involves the development, integration, and evaluation of ideas. Oates and Johnson (2025) also show that students' evaluative abilities improve when they are asked to criticize ChatGPT-generated essays using scientific sources. These findings are reinforced by Yusuf et al. (2024), who developed a gradual framework in the form of identification, conceptualization, investigation, evaluation, and synthesis of AI-generated information. Therefore, the use of reflective can be understood as the activity of checking accuracy, comparing perspectives, identifying weaknesses in arguments, and independently revising GenAI's answers.

Dependency and Decreased Cognitive Engagement

GenAI can also hinder the development of critical thinking if students rely too much on instant answers. Firdaus et al. (2025) found that AI provides efficiency in completing tasks, but over-reliance can weaken students' analytical, critical, and creative abilities. Baiquni et al. (2025) identified a decline in self-literacy, technology dependence, and verification needs due to ChatGPT's output not always matching authoritative sources. Rahmanto et al. (2024) also concluded that the integration of AI without pedagogical strategies can limit creativity and reduce critical thinking. This suggests that the intensity of GenAI use does not always result in a positive impact because its effect depends on the level of control, verification, and cognitive engagement of students.

Student Perception and Implementation Context

Students show non-uniform perceptions of the influence of AI on thinking skills. Veronica et al. (2025) found that some students consider AI to improve analytical skills, while others consider it to reduce these skills. These differences in perception show that GenAI's impact is influenced by the purpose of use, AI literacy, academic integrity, assignment design, and lecturer direction. Students who use GenAI to generate final answers tend to have lower cognitive engagement than students who use it to obtain feedback, test arguments, and

find alternative perspectives. Therefore, GenAI integration needs to be accompanied by ethical guidelines and learning activities that require students to explain, verify, and account for their answers.

Research Synthesis and Gaps

The literature shows that the relationship between GenAI use and critical thinking is conditional: active, reflective, and evaluative use has the potential to strengthen analytical skills, while passive use can lead to dependence. However, most Indonesian studies use a limited sample, measure perceptions at a time, or focus on the context of a particular course. International experimental research also generally evaluates interventions in the short term. These limitations support the need for longitudinal correlational research that measures the use of GenAI and changes in students' critical thinking skills at the beginning and end of the semester. This approach allows researchers to distinguish the intensity of general use from reflective use while assessing its development over time.

Research Hypothesis Development

Previous studies have shown that the use of GenAI can support the development of critical thinking skills if students use it to explore ideas, compare arguments, test the accuracy of information, and revise their thinking results. Rizaldi et al. (2024), Ria and Yuda (2025), and Rahmawati et al. (2025) found that the targeted use of ChatGPT is related to improving analytical skills, problem solving, and more in-depth question formulation. Nasr et al. (2025) also emphasized that active collaboration between humans and AI has more potential to develop critical thinking skills than the passive use of AI. Based on these findings, the more reflective and evaluative the use of GenAI by students, the more likely it is that there will be an increase in critical thinking skills during the learning process. Therefore, the main hypotheses of the study are formulated as follows:

H1 : The use of Generative Artificial Intelligence is positively and significantly related to the development of students' critical thinking skills in higher education.

However, the intensity of GenAI use does not always result in the development of critical thinking if students only receive or copy the answers produced. Firdaus et al. (2025), Baiquni et al. (2025), and Rahmanto et al. (2024) show that reliance on AI can reduce students' cognitive engagement, analytical ability, and independence. This shows that the quality of use is more important than the frequency of use alone. Reflective use involves verifying sources, evaluating arguments, comparing perspectives, and revising AI outputs, while passive use tends to transfer thought processes to technology. Based on these arguments, the following hypothesis is formulated:

H2 : The reflective and evaluative use of GenAI was positively and significantly related to the increase in students' critical thinking scores from the beginning to the end of the semester.

H3 : The passive and dependent use of GenAI is negatively and significantly related to the development of students' critical thinking skills.

Because the research was conducted longitudinally, the development of critical thinking was shown through changes in test scores between the beginning and end of the semester. The integration of GenAI in learning accompanied by analysis, verification, and reflection activities is expected to improve students' ability to evaluate information and draw conclusions based on evidence. Thus, the difference hypothesis is formulated as follows:

H4 : There was a significant increase in students' critical thinking ability scores between the initial measurement and the end of the semester measurement.

The zero hypothesis used in statistical testing is as follows:

H0₁ : The use of Generative Artificial Intelligence is not significantly related to the development of students' critical thinking skills.

H0₂ : There was no significant difference between students' critical thinking ability scores at the beginning and end of the semester.

The relationship between variables and the direction of the research hypothesis is presented in a conceptual framework in Figure 1.

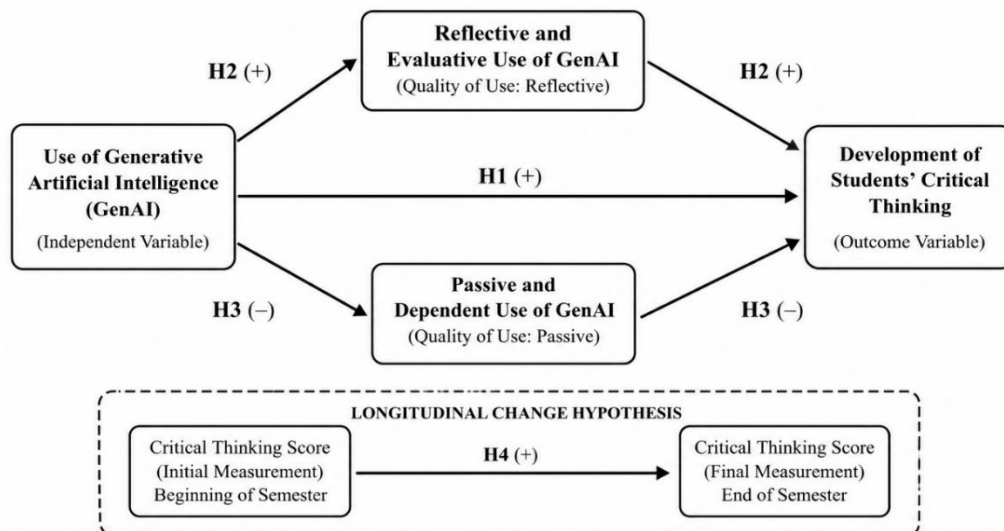


Figure 1. Framework

METHODOLOGY

Types of Research

This study uses a quantitative approach with a longitudinal correlational design. The correlational design was used to test the direction and strength of the relationship between the use of Generative Artificial Intelligence (GenAI) as an independent variable and the development of critical thinking skills as a dependent variable. The study did not provide experimental treatment and was therefore not intended to conclude a direct cause-and-effect relationship.

Measurements are carried out twice in one semester, namely at the beginning of the semester (time 1/T1) and the end of the semester (time 2/T2). The longitudinal design was chosen so that changes in critical thinking ability can be observed more accurately than the cross-sectional design that only

performs a one-time measurement (Hutapea & Kabatiah, 2025; Nazila et al., 2025).

Population and Sample

The research population consists of 1,238 active undergraduate students from three universities in East Java in the 2025/2026 academic year. The population includes 480 students from college A, 405 students from college B, and 353 students from college C. The three institutions were chosen because they have allowed the use of GenAI in learning while still practicing academic integrity.

Respondents must be active students in semesters 2–8, attend lectures for one semester, have used GenAI for academic purposes in the last three months, and are willing to take two stages of measurement. Students who do not complete one of the measurements or fill in the instrument incompletely are excluded from the analysis.

The sample was determined through proportionate stratified random sampling based on universities, semesters, and clusters of fields of science. This probability technique is used so that each group of students is proportionally represented and reduces the dominance of certain groups in the sample (Hutapea & Kabatiah, 2025).

Calculations using G*Power 3.1 showed the need for a minimum of 164 respondents with an effect size of 0.25, a significance level of 0.05, and a statistical strength of 0.90. A total of 210 students took part in the initial measurement, but only 186 students completed both stages, with a retention rate of 88.6%. The final sample consisted of 72 students from college A, 61 from college B, and 53 from college C.

Research Instruments

Data were collected using the GenAI Usage Questionnaire, the Critical Thinking Ability Test, and the respondent characteristics form. Characteristic forms are used to identify age, gender, semester, field of study, college, experience using GenAI, and the most frequently used applications.

The GenAI Usage Questionnaire consists of 24 items with a five-level Likert scale. The instrument measured three dimensions, namely the intensity of use as many as 6 items, the reflective and evaluative use of 10 items, and the passive use and dependence of 8 items.

The questionnaire items were developed from the Generative AI Literacy for Learning Scale which includes prompt skills, self-paced learning, and critical thinking (Gümüř & Kara, 2025). Indicators of benefits, usage patterns, and concerns about GenAI also refer to Chan and Hu (2023), then adjusted to the context of Indonesian students based on Agustin et al. (2024), Hutapea and Rantung (2024), and Manurung et al. (2023).

The validity of the content was assessed by five experts and obtained an Aiken index of 0.84–0.96. A trial of 60 students showed a CFI score = 0.936, TLI = 0.927, RMSEA = 0.058, and SRMR = 0.052. Cronbach's alpha value of 0.90 indicates that the questionnaire has good reliability.

Critical thinking skills were measured using case-based tests in two parallel forms. Form A is used in the initial measurement, while Form B is used

in the final measurement. Each form contains 25 questions that measure interpretation, analysis, evaluation, inference, and explanation, with a score range of 0–100.

Test development refers to the ability to identify problems, examine information, compare alternatives, and draw conclusions based on evidence (Amin et al., 2023; Darwin et al., 2024). The validity of the content obtained an Aiken index of 0.86–0.97. The reliability of the KR-20 was 0.84 for Form A and 0.86 for Form B, while the correlation between the two forms of the test was 0.81.

Data Collection Procedure

The study was carried out for 16 weeks. The initial stage includes instrument preparation, expert validation, trials, item revisions, research licensing, and ethical approval. Respondents obtained information regarding the purpose of the research, the confidentiality of the data, the voluntary nature of participation, and the right to withdraw.

In the first and second weeks, respondents filled out the characteristics form, the GenAI Use Questionnaire, and the Form A Critical Thinking Test.

During the semester, students participated in regular lecture activities without special intervention from researchers. The use of GenAI occurs naturally for activities such as searching for ideas, understanding material, composing writing, analyzing information, and completing academic tasks (Agustin et al., 2024; Hasim et al., 2023).

In the 15th and 16th weeks, respondents again filled out a questionnaire and took the Form B Critical Thinking Test. Of the 210 initial participants, as many as 186 students completed all stages and were included in the analysis.

Data Analysis Techniques

The data were analyzed using IBM SPSS Statistics 29, while the validity of the construct was tested with IBM SPSS AMOS 29. G*Power 3.1 is used to determine the requirements of the sample. The initial analysis included checking for missing data, input errors, extreme values, normality, linearity, homokedasticity, and multicollinearity.

Descriptive statistics are used to present frequencies, percentages, averages, standard deviations, minimum values, and maximums. The reliability of the instrument was tested using Cronbach's alpha for the questionnaire and the KR-20 for the critical thinking test.

The paired-samples t-test is used to compare critical thinking scores at the beginning and end of the semester and test H4. The magnitude of the change is reported using Cohen's d. The progression score is calculated by subtracting the initial score from the final score.

Pearson correlation was used to test the relationship between GenAI use and changes in critical thinking scores as an H1 test. Multiple linear regression was used to test the contribution of reflective use, passive use, and intensity of GenAI to critical thinking development as an H2 and H3 test.

Initial scores of critical thinking, college, semester, and gender were included as control variables. All tests used a significance level of 0.05 and were carried out in two directions. The determination coefficient is used to determine the proportion of critical thinking development that can be explained by the use of GenAI.

RESEARCH RESULT

Respondent Characteristics

A total of 210 students took part in the initial measurement, while 186 students completed the initial and final measurements so that the research retention rate reached 88.6%. A total of 24 students were not included in the analysis because they did not take the final measurement or fill out the instrument incompletely. Comparison of preliminary data showed that there was no significant difference between respondents who completed the study and those who exited in GenAI usage scores and initial critical thinking skills ($p > 0.05$).

Respondents came from three universities in East Java. The majority of respondents were women as many as 112 students or 60.2%, while men amounted to 74 students or 39.8%. The average age of the respondents was 20.7 years with a standard deviation of 1.4 years and an age range of 18–24 years.

Table 1. Characteristics of respondents

Features	Category	n	%
Colleges	College A	72	38,7
	College B	61	32,8
	College C	53	28,5
Gender	Women	112	60,2
	Male	74	39,8
Semester	Semester 2–4	96	51,6
	Semester 5–8	90	48,4
Cluster of sciences	Education	74	39,8
	Social and humanities	62	33,3
	Science and technology	50	26,9
Total		186	100,0

Generative Artificial Intelligence Use Patterns

ChatGPT is the most frequently used GenAI application by respondents, which is as many as 135 students or 72.6%. Other applications used were Gemini as many as 31 students or 16.7%, Perplexity as many as 12 students or 6.5%, and Microsoft Copilot as many as 8 students or 4.3%.

The use of GenAI is most often aimed at understanding lecture materials, developing ideas, compiling writing, and finding alternative task completions. Because respondents can choose more than one use destination, the percentage of usage does not amount to 100%.

Table 2. Academic objectives of GenAI use

Intended use	n	%
Understand or explain material	146	78,5
Develop writing ideas and frameworks	134	72,0
Assist with the preparation of written assignments	125	67,2
Compare alternative answers	111	59,7

Search for initial terms or information	102	54,8
Checking the language and structure of the writing	98	52,7
Helps with data analysis	61	32,8

At the end of the semester, the average score of GenAI's usage intensity was 3.72 on a scale of 1–5. Reflective and evaluative use obtained the highest average of 3.81, while passive and dependency use obtained an average of 2.48. These results show that there is a variation in the pattern of GenAI use among students.

Table 3. Descriptive statistics of GenAI usage

Dimensions	Average	SD	Minimum	Maximum
Intensity of use	3,72	0,61	1,83	5,00
Reflective and evaluative use	3,81	0,58	2,00	5,00
Passive use and dependency	2,48	0,67	1,00	4,63
Overall GenAI Usage	3,65	0,52	2,13	4,92

Instrument Quality and Assumption Checks

The GenAI usage questionnaire showed good internal consistency with Cronbach's alpha value of 0.91. The reliability value of each dimension was 0.82 for intensity of use, 0.89 for reflective and evaluative use, and 0.86 for passive use and dependence. The critical thinking test also showed good reliability, with a KR-20 value of 0.85 at the initial measurement and 0.87 at the final measurement.

Preliminary examination shows that the data meet the assumptions for parametric analysis. The *values of skewness* and *kurtosis* of the entire variable were in the range –0.63 to 0.51. Examination of the residual graph shows a relationship that is linear and the residual variance is relatively constant. The value of *the variance inflation factor* is in the range of 1.12–1.71 so that there is no problem of multicollinearity. A Durbin–Watson value of 1.94 indicates the absence of residual autocorrelation, while a maximum Cook's distance value of 0.18 indicates that there are no observations that significantly affect the regression model.

Changes in Critical Thinking Skills

The average score of critical thinking skills increased from 68.4 at the beginning of the semester to 74.9 at the end of the semester. Thus, the average score increase is 6.5 points. Improvements were found in all dimensions of critical thinking, with the highest increase in the evaluation dimension of 7.6 points and analysis of 7.1 points.

Table 4. Changes in critical thinking skills based on dimensions

Dimensions	Beginning of the semester	End of semester	Changes
Interpretasi	70,2	76,1	5,9
Analysis	67,5	74,6	7,1
Evaluation	65,8	73,4	7,6
Inference	69,1	75,0	5,9
Explanation	69,4	75,4	6,0
Overall score	68,4	74,9	6,5

A total of 137 students or 73.7% experienced an increase in critical thinking scores. A total of 28 students or 15.1% showed relatively stable changes, while 21 students or 11.3% experienced a decrease in scores.

The results of the *paired-samples t-test* showed that the difference in scores at the beginning and end of the semester was statistically significant, $t(185) = 10.81$; $p < 0.001$. The 95% confidence interval for the score improvement is in the range of 5.31–7.69. Cohen's *value of d* of 0.79 indicates a medium to large effect size.

Table 5. Results of the critical thinking score difference test

Measurement	Average	SD	Average difference	t	p	Cohen's d
Beginning of the semester	68,4	8,6				
End of semester	74,9	8,2	6,5	10,81	<0.001	0,79

No significant difference was found in the improvement in scores by college, $F(2.183) = 1.18$; $p = 0.309$. The average increase in A college was 6.9 points, B college was 6.2 points, and C college was 6.3 points.

The Relationship between the Use of GenAI and the Development of Critical Thinking

The critical thinking skill development score is calculated by subtracting the initial score from the final score. Pearson's correlation results showed that the overall use of GenAI was positively and significantly related to the development of critical thinking skills, $r = 0.34$; $p < 0.001$.

The reflective and evaluative use of GenAI showed a moderate positive relationship with the development of critical thinking, $r = 0.46$; $p < 0.001$. This means that students who examine information more frequently, compare arguments, evaluate outputs, and revise GenAI answers tend to show a greater increase in critical thinking scores.

In contrast, passive use and dependence were negatively associated with the development of critical thinking, $r = -0.30$; $p < 0.001$. The intensity of use showed a weak positive relationship, although it remained statistically significant, $r = 0.19$; $p = 0.009$.

Table 6. Correlation of the use of GenAI with the development of critical thinking

GenAI usage variables	r	p	Direction of contact
Overall GenAI Usage	0,34	<0.001	Positive
Intensity of use	0,19	0,009	Weak positives
Reflective and evaluative	0,46	<0.001	Moderate positive
Passivity and dependency	-0,30	<0.001	Negatives

Contribution of GenAI Use Patterns

Multiple linear regression was used to test the contribution of use intensity, reflective and evaluative use, and passive use and dependence on

changes in critical thinking scores. Early critical thinking scores, gender, semester, and college were included as control variables.

The regression model was statistically significant, $F(8.177) = 11.40$; $p < 0.001$, with $R^2 = 0.34$ and adjusted $R^2 = 0.31$. The model explains the 34% variation in the development of students' critical thinking skills.

Reflective and evaluative use were the strongest positive predictors, with $B = 4.15$; $\beta = 0.38$; $p < 0.001$. Each one-point increase in reflective and evaluative use was associated with an approximately 4.15-point increase in critical thinking developmental scores after the other variables were controlled.

Passive use and dependency were significant negative predictors, with $B = -2.31$; $\beta = -0.22$; $p = 0.001$. Meanwhile, the intensity of use no longer made a significant contribution after reflective and passive usage patterns were incorporated into the model, $B = 0.73$; $\beta = 0.08$; $p = 0.210$.

Table 7. Multiple linear regression results

Predictor	B	SE	b	t	p
Intensity of GenAI use	0,73	0,58	0,08	1,26	0,210
Reflective and evaluative use	4,15	0,76	0,38	5,46	<0.001
Passive use and dependency	-2,31	0,71	-0,22	-3,25	0,001
Early critical thinking score	-0,18	0,05	-0,26	-3,79	<0.001
Semester	0,19	0,21	0,06	0,91	0,365
Gender	0,37	0,63	0,04	0,59	0,557
College B	-0,51	0,72	-0,05	-0,71	0,479
College C	-0,76	0,75	-0,07	-1,01	0,314

Remarks: College A is used as a reference category; The dependent variable is a change in critical thinking score.

Hypothesis Testing Results

The test results show that all research hypotheses are supported, except that the use of GenAI is only measured based on intensity after the usage pattern variables are controlled. Reflective use provides positive relationships and contributions, while passive use provides negative relationships and contributions.

Table 8. Hypothesis testing summary

Hipotesis	Statement	Results
H1	The use of GenAI is positively related to the development of critical thinking	Supported
H2	Reflective and evaluative use is positively associated with increased critical thinking	Supported
H3	Passive use and dependence are negatively related to the development of critical thinking	Supported
H4	There was an increase in critical thinking scores between the beginning and end of the semester	Supported

Based on these results, H_{01} which stated that there was no relationship between the use of GenAI and the development of critical thinking was rejected. H_{02} which stated that there was no difference in critical thinking score between the beginning and end of the semester was also rejected.

DISCUSSION

The results showed that students' critical thinking skills increased from an average score of 68.4 at the beginning of the semester to 74.9 at the end of the semester. The difference of 6.5 points is statistically significant, $t(185) = 10.81$; $p < 0.001$, with a measure of Cohen's effect d of 0.79. The biggest improvement occurred in the evaluation and analysis dimensions, which showed the development of students' ability to assess the accuracy of information, compare arguments, and draw conclusions based on evidence. These findings are in line with Rizaldi et al. (2024) and Ria and Yuda (2025), who show that the targeted use of ChatGPT can support students' analytical and problem-solving skills.

The reflective and evaluative use of GenAI has a moderate positive relationship with the development of critical thinking, namely $r = 0.46$; $p < 0.001$. The regression results also showed that reflective use was the strongest positive predictor with $\beta = 0.38$; $p < 0.001$. College students who use GenAI to compare answers, examine sources, evaluate arguments, and revise AI outputs tend to experience greater critical thinking improvements. These findings support the concept of GenAI as a cognitive scaffold, which is a tool that helps the thinking process without replacing students' intellectual control. These results are in line with Nasr et al. (2025) and Essel et al. (2024), who affirm that active human-AI collaboration is more beneficial than passive use of AI.

On the other hand, the passive use of GenAI and its dependence was negatively related to the development of critical thinking, namely $r = -0.30$; $p < 0.001$. In the regression model, passive use remained a significant negative predictor with $\beta = -0.22$; $p = 0.001$. This finding can be explained through the concept of cognitive offloading, which is the transfer of the process of analysis, evaluation, and decision-making to technology. When students simply receive or copy AI answers, they do not undergo the cognitive processes necessary to develop critical thinking skills. These results support the findings of Gerlich (2025), Firdaus et al. (2025), and Baiquni et al. (2025) regarding the risk of dependence, decreased self-literacy, and reduced cognitive engagement due to uncontrolled use of AI.

The intensity of GenAI use only showed a weak positive relationship with the development of critical thinking, namely $r = 0.19$; $p = 0.009$. The effect becomes insignificant after reflective and passive use is incorporated into the regression model. This shows that the quality of use is more decisive than the frequency of use. Students who often use GenAI do not necessarily get critical thinking development if their use is only oriented towards the final answer. Conversely, less frequent use can still provide benefits when it involves verification, evaluation, and reflection. These findings clarify the results of Putri and Panduwinata (2025) and Rahmanto et al. (2024) that the impact of AI depends on the strategy of its use in learning.

The research model explained 34% variation in the development of critical thinking skills, while the rest could be influenced by other factors, such as early ability, learning motivation, digital literacy, metacognition, assignment design, quality of lecturer feedback, and academic environment. Early critical thinking scores were negatively associated with improved scores, suggesting that students with lower initial scores had greater room for development. Meanwhile, gender, semester, and college had no significant effect. These results indicate that the benefits and risks of using GenAI can arise in different groups of students, so that the ability to use AI critically is not only determined by academic experience.

Theoretically, this study shows that GenAI can function as a cognitive scaffold as well as a source of cognitive offloading. Positive impacts arise when students remain in control of the thinking process, while negative impacts occur when the process is left to technology. Practically, colleges need to develop learning that requires students to examine sources, identify AI errors, compare answers, explain the revision process, and account for the results of their work. Assessment also needs to be process-oriented through case-based assignments, source criticism, presentations, and reflective journaling, rather than just assessing the final product.

This study has limitations because it uses a correlational design without a control group, so the increase in critical thinking cannot be fully considered a result of the use of GenAI. Data on the use of AI was also obtained through questionnaires that have the potential to experience perception bias. In addition, the measurement was only carried out twice and respondents were limited to three universities in East Java. Further research is suggested using experimental or quasi-experimental designs, taking measurements more than twice, analyzing the history of AI prompts and outputs, and involving universities from more diverse regions and disciplines. Variables such as AI literacy, metacognition, motivation, trust in AI, and cognitive offloading also need to be analyzed as mediators or moderators.

CONCLUSIONS AND RECOMMENDATIONS

This study concluded that students' critical thinking skills increased significantly from an average score of 68.4 to 74.9 during one semester. Reflective and evaluative use of GenAI had a moderate positive relationship with critical thinking development, while passive use and dependency showed a negative relationship. These findings confirm that the benefits of GenAI are not primarily determined by the frequency of its use, but by the quality of student interaction in examining sources, comparing arguments, evaluating outputs, and revising answers. Thus, GenAI can function as a cognitive scaffold when used actively, but can encourage cognitive offloading when it replaces students' thinking processes.

Colleges are advised to integrate GenAI through guidelines that emphasize information verification, transparency of use, reflection, and academic integrity. Lecturers need to design case-based assignments, source criticism, presentations, as well as prompt documentation and revision processes so that assessments don't just focus on the final product. Students also need to be

equipped with AI literacy to be able to recognize biases, inaccuracies, and limitations of GenAI's outputs so that technology remains a learning tool, not a substitute for intellectual independence.

ADVANCED RESEARCH

Further research is recommended using experimental or quasi-experimental designs with control groups, longer observation periods, and more than two waves of measurements. Follow-up studies also need to involve universities from more diverse regions and disciplines, using prompt or digital activity history data, and examining the role of AI literacy, metacognition, motivation, trust in AI, and cognitive offloading as mediation or moderation variables.

ACKNOWLEDGMENT

The author expressed his gratitude to all students and universities who have participated in this research. Appreciation was also given to experts, lecturers, and institutions who supported the process of instrument validation, data collection, and research completion.

REFERENCES

- Agustin, F. A. P., Wulandari, O., & Rusydiyah, E. F. (2024). Enhancing digital literacy through Chat GPT: The impact on Islamic education research skills among students at State Islamic University. *Tribakti: Jurnal Pemikiran Keislaman*, 35(2), 181–198. <https://doi.org/10.33367/tribakti.v35i2.5391>
- Akastangga, M. D. F., Harmonis, S., & Al Hafidz, R. A. (2023). The impact of ChatGPT on the critical thinking ability of UIN Sunan Kalijaga students. *Matrix: Jurnal Manajemen Teknologi dan Informatika*, 13(3), 157–165. <https://doi.org/10.31940/matrix.v13i3.157-165>
- Amin, A. M., Adiansyah, R., & Hujjatusnaini, N. (2023). The contribution of communication and digital literacy skills to critical thinking. *Jurnal Pendidikan Sains Indonesia*, 11(3), 697–712. <https://doi.org/10.24815/jpsi.v11i3.30838>
- Baiquni, A., Rusli, T. I., Juhardin, J., Kholida, L., & Arsam, A. (2025). Analysis of the impact of ChatGPT on learning productivity of Islamic education students. *EDURELIGIA: Jurnal Pendidikan Agama Islam*, 9(1), 68–81. <https://doi.org/10.33650/edureligia.v9i1.10407>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024).

- Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, 11(1), Article 2290342. <https://doi.org/10.1080/2331186X.2023.2290342>
- 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, Article 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
- Fadholi, M. A. F. (2025). Application of the Technology Acceptance Model to Chat GPT to measure the critical thinking power of UIN Sunan Ampel Surabaya students. *Multitek Indonesia: Jurnal Ilmiah*, 19(1), 85–96. <https://doi.org/10.24269/mtkind.v19i1.12209>
- Firdaus, J. A., Ummah, R. I., Aprialini, R. R., Fithriyyah, A., Mahsusi, & Faizin, A. (2025). Ketergantungan penggunaan kecerdasan buatan (AI) pada tugas akademik mahasiswa terhadap kemampuan berpikir kritis dan kreatif. *Didaktika: Jurnal Kependidikan*, 14(1), 1203–1214. <https://doi.org/10.58230/27454312.1634>
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), Article 6. <https://doi.org/10.3390/soc15010006>
- Gümüş, M. M., & Kara, M. (2025). Development and validation of the Generative AI Literacy for Learning Scale (GenAI-LLs). *Australasian Journal of Educational Technology*, 41(4), 1–16. <https://doi.org/10.14742/ajet.10236>
- Hasim, S., Khaira, M., Karunia, G. C. M. K., Limbong, J. A., & Amelia, D. (2023). Pengaruh penggunaan Chat GPT terhadap minat baca mahasiswa. *EDUTECH: Jurnal Teknologi Pendidikan*, 22(3), 269–279. <https://doi.org/10.17509/e.v22i3.62944>
- Hutapea, G. L., & Rantung, J. (2024). Hubungan persepsi mahasiswa keperawatan dengan sikap penggunaan Chat GPT. *Jurnal Penelitian Inovatif*, 4(2), 519–528. <https://doi.org/10.54082/jupin.350>
- Hutapea, N. M., & Kabatiah, M. (2025). Pengaruh penggunaan ChatGPT terhadap kemampuan berpikir kritis mahasiswa Jurusan PPKN Angkatan 2023 FIS Unimed. *SOCIAL: Jurnal Inovasi Pendidikan IPS*, 5(3), 1062–1071. <https://doi.org/10.51878/social.v5i3.6709>
- Jasmadi, J., Saddam, S., & Lasri, L. (2024). Critical thinking: An analytical study on the impact of artificial intelligence (AI) usage on students at Al Washliyah Darussalam University, Banda Aceh. *Sekumpul: Journal of Multidisciplinary Education Sciences*, 2(1), 25–32. <https://doi.org/10.62568/jomes.v2i1.168>

- Lukman, L., Agustina, R., & Aisy, R. (2023). Problematika penggunaan artificial intelligence (AI) untuk pembelajaran di kalangan mahasiswa STIT Pemalang. *Madaniyah*, 13(2), 242–255. <https://doi.org/10.58410/madaniyah.v13i2.826>
- Manurung, O., Destiani, A. C., Sugiarto, J., Lolo, A. T. A., & Chai, K. (2023). Identifikasi pengaruh penggunaan ChatGPT terhadap kemampuan berfikir mahasiswa di Universitas Atma Jaya Yogyakarta Prodi Sistem Informasi Angkatan 2021. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 3(2), 342–352. <https://doi.org/10.24002/konstelasi.v3i2.7241>
- Nasr, N. R., Tu, C.-H., Werner, J., Bauer, T., Yen, C.-J., & Sujo-Montes, L. (2025). Exploring the impact of generative AI ChatGPT on critical thinking in higher education: Passive AI-directed use or human–AI supported collaboration? *Education Sciences*, 15(9), Article 1198. <https://doi.org/10.3390/educsci15091198>
- Nazila, M., Mahardika, D., & Sugara, H. (2025). Pengaruh penggunaan ChatGPT terhadap kemampuan berpikir kritis pada mahasiswa bimbingan dan konseling di Universitas Indraprasta PGRI. *TheraEdu: Journal of Therapy and Educational Psychology*, 1(1), 11–20. <https://doi.org/10.63203/021817500>
- Oates, A., & Johnson, D. (2025). ChatGPT in the classroom: Evaluating its role in fostering critical evaluation skills. *International Journal of Artificial Intelligence in Education*, 35(4), 1793–1824. <https://doi.org/10.1007/s40593-024-00452-8>
- Putri, M. A., & Panduwinata, L. F. (2025). Pengaruh penggunaan artificial intelligence terhadap berpikir kritis mahasiswa. *Jurnal Kajian Ilmu Pendidikan*, 6(2), 605–616. <https://doi.org/10.55583/jkip.v6i2.1413>
- Rahmanto, A. A., Arum, M., Rahmawati, D. R., Wijayanti, V. C., & Ramadhan, G. (2024). Artificial intelligence dan critical thinking: Systematic literature review. *Jurnal PIPSI: Jurnal Pendidikan IPS Indonesia*, 9(3), 242–251. <https://doi.org/10.26737/jpipsi.v9i3.5545>
- Rahmawati, L., Yuliawati, R., Anisah, A., & Hermansyah. (2025). Analisis pengaruh pemanfaatan ChatGPT terhadap keterampilan menyusun pertanyaan efektif pada mahasiswa pendidikan ekonomi. *JATI: Jurnal Mahasiswa Teknik Informatika*, 9(5), 7858–7864. <https://doi.org/10.36040/jati.v9i5.14921>
- Refaldi, D. A., Faiz, A., Jawakory, M. R., & Rakhmawati, N. A. (2024). Analisis korelasi Pearson faktor pengaruh generative AI terhadap kemampuan berpikir kritis mahasiswa ITS Surabaya. *Jurnal Sistem Informasi dan Aplikasi*, 2(2), 9–19. <https://doi.org/10.52958/jsia.v2i2.7974>
- Ria, R. R. P., & Yuda, L. S. (2025). Efek Chat Generative Pre-training Transformer

- terhadap kemampuan berpikir kritis dan computational thinking mahasiswa dalam penulisan karya ilmiah di era Revolusi Industri 5.0. *UPGRADE: Jurnal Pendidikan Teknologi Informasi*, 2(2), 9–19. <https://doi.org/10.30812/upgrade.v2i2.4777>
- Rizaldi, M., Sasalia, P., & Pancarita. (2024). Efektivitas ChatGPT untuk mendorong berpikir kritis mahasiswa dalam menyelesaikan masalah aljabar. *EDU-MAT: Jurnal Pendidikan Matematika*, 12(1), 168–181. <https://doi.org/10.20527/edumat.v12i1.18668>
- Saraswati, A. R., Karmina, V. A., Efendi, M. P., Candrakanti, Z., & Rakhmawati, N. A. (2023). Analisis pengaruh ChatGPT terhadap tingkat kemalasan berpikir mahasiswa ITS dalam proses pengerjaan tugas. *Jurnal Pendidikan, Bahasa dan Budaya*, 2(4), 40–48. <https://doi.org/10.55606/jpbb.v2i4.2223>
- van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, 13(10), Article 998. <https://doi.org/10.3390/educsci13100998>
- Veronica, S. Y., Nurfairuzah, D. S., Aulia, I. R., & Azizy, P. Y. (2025). Persepsi mahasiswa dalam penggunaan AI terhadap keterampilan berpikir. *BASINDO: Jurnal Kajian Bahasa, Sastra Indonesia, dan Pembelajarannya*, 9(2), 328–341. <https://doi.org/10.17977/um007v9i22025p328-341>
- Yusuf, A., Bello, S., Pervin, N., & Tukur, A. K. (2024). Implementing a proposed framework for enhancing critical thinking skills in synthesizing AI-generated texts. *Thinking Skills and Creativity*, 53, Article 101619. <https://doi.org/10.1016/j.tsc.2024.101619>