Machine Learning for Predictive Evaluation of Students' Interactions with AI-Generated Content and Their Critical Thinking Levels*

Nataliia Dziubanovska[1,*,†], Vadym Maslii [1,†]

[1] West Ukrainian National University, Lvivska 11, 46009 Ternopil, Ukraine

Abstract

This article presents the results of an empirical study examining how the use of generative artificial intelligence models in the learning process relates to the level of critical thinking among economics majors. Based on a survey of 210 higher-education students, we collected data on the frequency of AI use, the nature of verification practices, and the structure of cognitive processes, the latter assessed using the standardized Watson–Glaser Critical Thinking Appraisal. Correlation analysis showed that critical thinking is most strongly associated with fact-checking of AI-generated content and a willingness to seek consultation when doubts arise. A binary logistic regression model confirmed the predictive significance of behavioral indicators for classifying students into a high–critical thinking group (AUC = 0.66). Paradoxically, higher levels of doubt unaccompanied by actual verification were associated with a lower likelihood of high test performance. We conclude that the key factor supporting critical thinking is not the mere use of AI, but the implementation of verification strategies during interaction with it. The findings can inform educational interventions aimed at fostering students' epistemic autonomy in digital environments.

Keywords

generative artificial intelligence; critical thinking; information verification; verification strategies; Watson–Glaser; logistic regression; student behavior; fact-checking

1. Introduction

In contemporary educational settings, generative artificial intelligence (AI) models are increasingly used by students as sources of information, ideas, and ready-made solutions. On the one hand, this creates new opportunities to accelerate learning; on the other, it heightens the risks of uncritical acceptance of automatically generated responses and diminished motivation for independent thinking. Under these conditions, the ability to verify the reliability of AI-generated content and to seek corroboration in authoritative sources becomes particularly important. Verification practices, in turn, may function as a factor that either supports – or, conversely, weakens – the development of critical thinking among higher-education students.

The relevance of this study stems from the need to empirically determine how students interact with AI-generated content and which behavioral strategies are most closely associated with high levels of critical thinking. To this end, we use data from a survey on AI-use practices in learning and results from the standardized Watson–Glaser Critical Thinking Appraisal. Subsequent application of machine learning methods enables the classification of student behavior types in the context of AI use and allows for a more precise assessment of how specific strategies for engaging with AI-generated content may influence the development of critical thinking.

*Applied Information Systems and Technologies in the Digital Society (AISTDS-2025), October 01, 2025, Kyiv, Ukraine

[1*] Corresponding author.
[†] These authors contributed equally.
n.dziubanovska@wunu.edu.ua (N. Dziubanovska); v.maslii@wunu.edu.ua (V. Maslii)
0000-0002-8441-5216 (N. Dziubanovska); 0000-0002-9672-9669 (V. Maslii)

2. Literature review

As L. F. Santos (2017) notes, critical thinking is an integral component of scientific practice and one of the key drivers of the development of science education [1]. B. Arisoy and B. Aybek (2021) emphasize that one of the fundamental aims of education is to cultivate critically and analytically minded individuals who can leverage acquired knowledge to improve their own quality of life and contribute to the advancement of society, culture, and civilization [2]. As T. Raj, R. Chauhan, R. Mehrotra, and M. Sharma (2022) observe [3], critical thinking is not merely the ability to reason in accordance with logical principles and the laws of rationality; it is also the capacity to apply these cognitive skills to inquiry and the solution of real-world problems. In contemporary higher education, students are expected to think critically while researching, evaluating, and interpreting information, thereby enabling the formulation of well-founded arguments and conclusions. H. Pervaiz, K. Ali, S. Razzaq, and M. Tariq (2025) underscore that higher education requires the development of critical thinking as a means of evaluating information, solving problems, and sustaining intellectual discourse [4].

The rise of artificial intelligence has sparked scholarly debate about its impact on the development of critical thinking. Both diametrically opposed theoretical positions and empirical quantitative findings can be observed. E. P. Ododo et al. (2024) [5] examined the challenges AI poses for cultivating critical thinking among university students. Using a structured questionnaire and descriptive statistics, including the independent-samples t test, the authors identified significant threats to critical thinking, including gender-based differences. Investigating AI's influence on the development of students' critical thinking, K. Szmyd and E. Mitera (2024) [6] conducted an online survey of 190 students at Polish universities; responses were grouped and visualized in charts, and the authors stress that these findings serve only as a starting point for further research. In a study by H. Pervaiz, K. Ali, S. Razzaq, and M. Tariq (2025) [4], students' perceptions of the effectiveness of AI tools in fostering their critical thinking were analyzed using Likert-scale questionnaires containing critical-type items, with both descriptive statistics and inferential methods; the results indicated a limited impact of AI on the development of critical thinking. In another study, A. M. Vieriu and G. Petrea (2025) [7], based on a sample of 85 second-year students at the National University of Science and Technology "Politehnica" (Bucharest), the perception, use, and effectiveness of AI tools were assessed; quantitative survey data were analyzed using frequency and percentage statistics, revealing overreliance on AI and diminished critical-thinking skills. Examining the impact of AI tools on critical thinking, M. Gerlich (2025) employed, in addition to descriptive statistics, analysis of variance, correlation analysis, multiple regression, and random forest regression, finding a significant negative correlation between the frequency of AI-tool use and critical-thinking ability [8]. H.-P. Lee et al. (2025), using logistic and linear regression, established that higher trust in AI is associated with lower levels of critical thinking, whereas greater self-confidence is associated with its strengthening. Qualitatively, AI is reshaping the nature of critical thinking by shifting emphasis toward information verification, integration of responses, and task management [9].

3. Methodology

The study's methodological approach is aimed at empirically assessing how the use of generative AI models for educational purposes may influence the development of students' critical thinking. The empirical base was formed from a survey of first- and second-year full-time bachelor's students enrolled in the Faculties of Economics and Management and Finance and Accounting at West Ukrainian National University (Ternopil). We focused on students in humanities-oriented economics programs, proceeding from the assumption that students in technical fields – owing to more intensive mathematical training – typically possess more advanced analytical competencies, which may serve as a natural factor supporting critical thinking [10]. In turn, we argue that economics students with a humanities profile warrant dedicated investigation into how interaction with generative AI affects the development of argumentative and evaluative cognitive skills.

In total, 210 respondents were surveyed. To collect empirical data, we designed a questionnaire comprising two parts: the internationally validated Watson–Glaser Critical Thinking Appraisal and an author-constructed survey capturing practices of applying generative AI models to academic tasks. Items were measured on a five-point Likert scale (from 1, "strongly disagree/not characteristic," to 5, "strongly agree/very characteristic"), enabling us to quantify not only the frequency of recourse to AI but also the intensity of initial verification of generated responses and the propensity to seek additional corroboration (from instructors or peers) when doubts arise (Table 1).

Table 1

Descriptive Statistics from the Student Survey on Practices of Interaction with AI

№	Question	1	2	3	4	5	Mean	Standard Deviation
		1 – "strongly disagree" / "don't feel," and 5 – "strongly agree" / "strongly feel"						
Q1	How often do you verify AI-provided information from other sources (e.g., refer to textbooks, websites, etc., for confirmation)?	13	33	78	62	22	3.23	1.04
Q2	When I receive information from AI, no matter how convincing it is, I always try to find confirmation in other sources.	8	38	87	43	32	3.25	1.05
Q3	If the information received contradicts my prior knowledge, I tend to thoroughly analyze it from different perspectives.	9	18	62	65	54	3.66	1.09
Q4	When I encounter unusual or ambiguous information, I usually make an effort to search for alternative explanations, even if the first one seems logical.	5	24	77	64	38	3.5	0.99
Q5	When using AI, I sometimes have doubts about the accuracy of the provided information and look for additional arguments to confirm or disprove it.	5	24	56	75	48	3.66	1.03
Q6	When the AI system provides me with an answer, I usually do not accept it immediately	9	36	72	55	36	3.35	1.09

	but analyze possible alternatives before using it in my work.						
Q7	When the information received from AI about academic subjects raises doubts, I feel the need to consult with teachers or discuss it with classmates.	20	47	74	37	30	3.05	1.17
Q8	Using AI helps me find solutions quickly, but sometimes I feel that it reduces my need to think independently.	11	24	59	61	53	3.58	1.14
Q9	How would you rate your ability to track hallucinations produced by AI?	31	40	81	47	9	2.82	1.08

In responding to the questionnaire, most participants indicated that they seek corroboration for AI-provided information from other sources (mean = 3.25, SD = 1.05). Students also report actively analyzing information when it conflicts with their prior knowledge (mean = 3.66, SD = 1.09), which supports their ability to evaluate data carefully from multiple perspectives.

Survey results further show that a majority attempt to find alternative explanations even when AI-generated content appears logical (mean = 3.50, SD = 0.99), evidencing cognitive flexibility. Respondents also express doubts regarding the accuracy of AI outputs and frequently look for additional arguments to either confirm or refute the provided information (mean = 3.66, SD = 1.03).

It is noteworthy that many students agree generative AI tools help them arrive at solutions more quickly, although this sometimes reduces the need for independent thinking (mean = 3.58, SD = 1.14). AI use is also viewed as contributing positively to fact-checking and the search for alternative viewpoints (mean = 3.44, SD = 1.03). However, the responses reveal difficulties in detecting AI "hallucinations" (mean = 2.82, SD = 1.08), indicating challenges in discerning reliability and heightened susceptibility to misinformation.

To measure students' critical-thinking proficiency, we used an adapted version of the Watson–Glaser test, widely regarded as an international standard for assessing the cognitive component of critical thinking. The maximum attainable score on the adapted version was 26 points, reflecting the sum of correct responses across task groups. The test content enables assessment of five key facets:

1. Drawing inferences – the ability to form logical conclusions based on facts and stated propositions;
2. Recognizing assumptions – the skill of identifying implicit statements that are not made explicit yet may influence how information is interpreted;
3. Deduction – evaluating the logical validity of conclusions proposed on the basis of given data;
4. Interpretation – the capacity to correctly construe factual material and make well-founded conclusions; and
5. Evaluating arguments – the ability to determine the strength of arguments, discerning compelling from weak lines of reasoning.

Accordingly, the total score reflected a composite level of students' cognitive maturity and their capacity for critical information processing.

To establish the internal consistency of the adapted Watson–Glaser test and to examine the structure of relationships among its components, we conducted a correlation analysis at the level of individual subtests. Pearson correlation coefficients were used to test the extent to which the five cognitive components of the test are associated with the overall critical-thinking score.

Additionally, correlation analysis was employed to identify associations between the overall critical-thinking score and behavioral variables characterizing students' interaction with generative AI (frequency of verification of AI outputs, degree of doubt, and willingness to seek consultation). This enabled us to determine the extent to which cognitive indicators of critical thinking are linked to actual verification strategies in instructional practice.

To build a predictive model classifying students by their interaction types with generative AI, we applied binary logistic regression with a target variable representing critical-thinking level: High CM (1) for students with a Watson–Glaser total ≥ 18 and Non-High CM (0) for all others (≤ 17). Predictors comprised behavioral indicators of AI interaction: the frequency of verifying AI-provided information, the degree of doubt about its accuracy, and the willingness to consult instructors/peers. Features were standardized; model training used L2 regularization to minimize overfitting. Predictor informativeness was assessed via model coefficients and odds ratios (OR), interpreted on the standardized scale. Model quality was evaluated using ROC–AUC. All computations were performed in Python.

4. Results

The findings indicate that most students use generative AI in their coursework, though the frequency of use varies: some rely on it regularly, while others do so only intermittently (Figure 1).

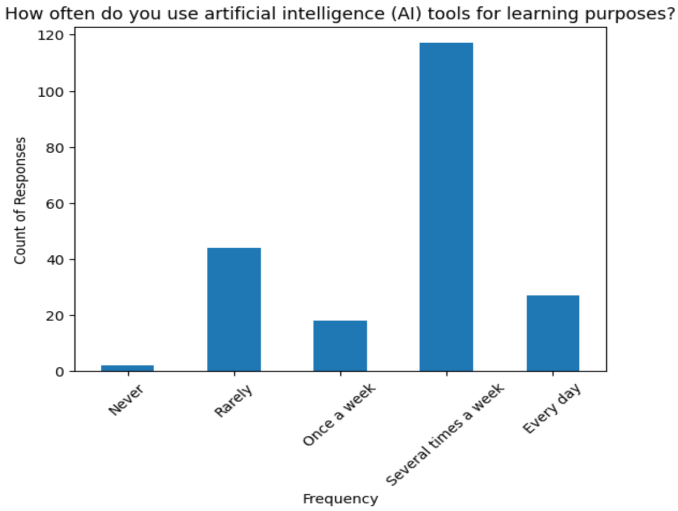

Figure 1: Frequency of AI Use in the Educational Process.

The most common response was "several times a week," selected by 117 respondents (55.7%). Twenty-seven students (12.8%) reported using generative AI for academic purposes daily, 18 (8.6%) once a week, and 44 (21.0%) only rarely; just 4 students (1.9%) stated they had never used AI in the learning process. Overall, this suggests that a substantial share of participants (68.5%) actively employ generative-AI tools in their studies.

According to the Watson–Glaser test results (mean = 15.93; SD = 2.47; minimum = 9 out of 26; maximum = 21 out of 26), the distribution of critical-thinking levels was as follows: 25.2% of students (53 individuals) scored in the low range (≤ 14), 49.0% (103 individuals) in the medium range (15–17), and 25.7% (54 individuals) in the high range (≥ 18) (Figure 2). These findings indicate that roughly half of the sample falls within an "optimal range," whereas about one in four students would benefit from additional measures to develop analytical and argumentation skills.

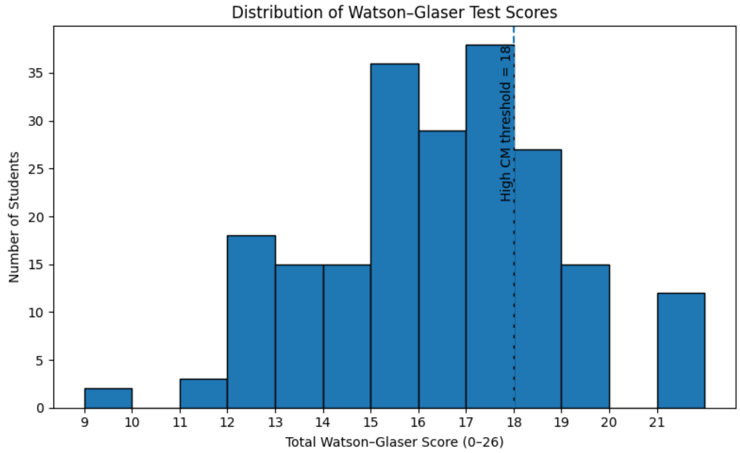

Figure 2: Distribution of Watson–Glaser Test Scores.

Correlation analysis confirmed that all five Watson–Glaser subtests (drawing inferences, recognizing assumptions, deduction, interpretation, and evaluating arguments) are significantly associated with the total score, with relationship strength ranging from moderate to high. The strongest correlation was observed for Evaluating Arguments ($r = 0.561$; $p < 0.001$), followed closely by Interpretation ($r = 0.535$; $p < 0.001$) and Recognizing Assumptions ($r = 0.519$; $p < 0.001$). The correlations for Deduction ($r = 0.414$; $p < 0.001$) and Drawing Inferences ($r = 0.359$; $p < 0.001$) remain moderate, yet still indicate a meaningful contribution to the overall critical-thinking score (Figure 3).

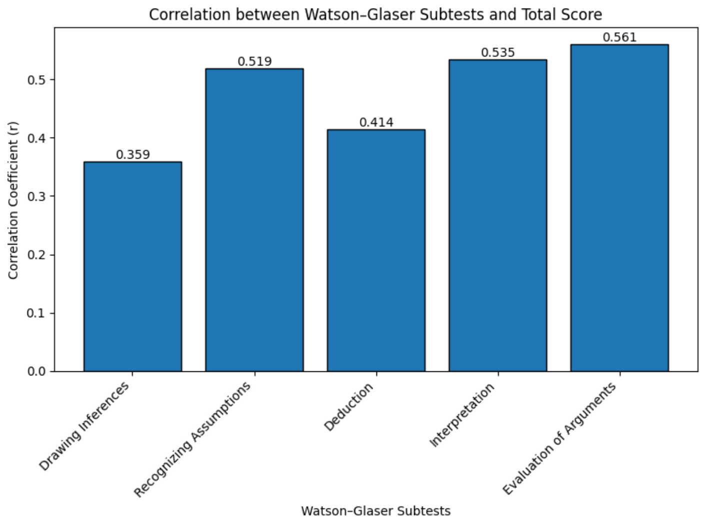

Figure 3: Correlation between Watson–Glaser Subtests and Total Score.

This hierarchy of coefficients suggests that skills in evaluating arguments and interpretation contribute most to the composite result, while drawing inferences and deduction play a somewhat smaller, though still important, role.

At the same time, a substantial share of students with high levels of critical thinking demonstrate the ability to apply verification strategies effectively when working with generative AI models, as evidenced by a strong correlation between the total Watson–Glaser score and the frequency of checking AI outputs ($r = 0.586$; $p < 0.001$). A moderate yet significant association also emerged with the tendency to consult instructors or peers when in doubt ($r = 0.385$; $p < 0.001$). The relationship with the degree of doubt regarding the accuracy of AI-generated responses was weak but statistically significant ($r = 0.276$; $p < 0.001$) (Figure 4).

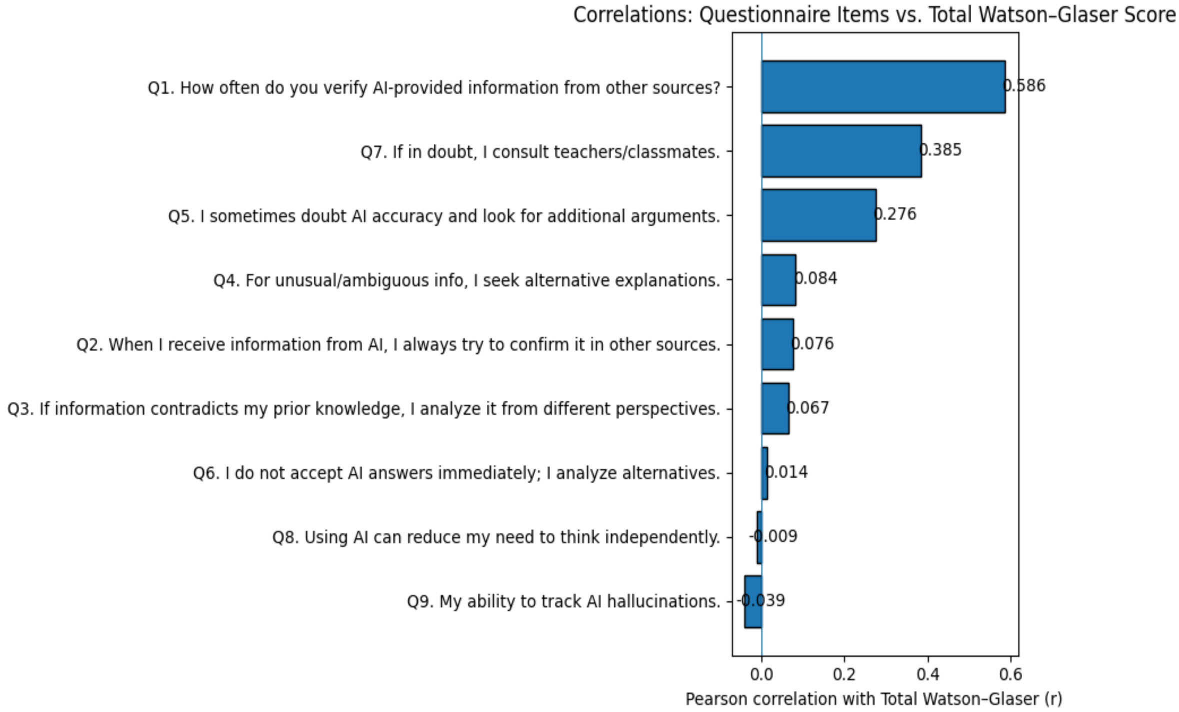

Figure 4: Correlations between Questionnaire Items and Total Watson–Glaser Score.

Guided by the correlation results, we selected three behavioral variables that capture the most salient verification-analytic practices in students' interaction with generative AI models, along with an integrative cognitive indicator – the total Watson–Glaser score.

These four variables – (i) frequency of content verification, (ii) degree of doubt regarding the accuracy of AI responses, (iii) willingness to consult instructors/peers, and (iv) the composite critical-thinking score – exhibited the clearest and statistically significant associations with critical-thinking level. The first three represent distinct behavioral strategies when working with AI (information checking, epistemic skepticism, and social verification via consultation), while the Watson–Glaser total reflects the realized cognitive capacity for critical evaluation of information.

This predictor set provides the most substantive input for constructing a binary predictive model (Figure 5) that estimates the likelihood a student belongs to the high critical-thinking group based on observed practices of interaction with generative AI.

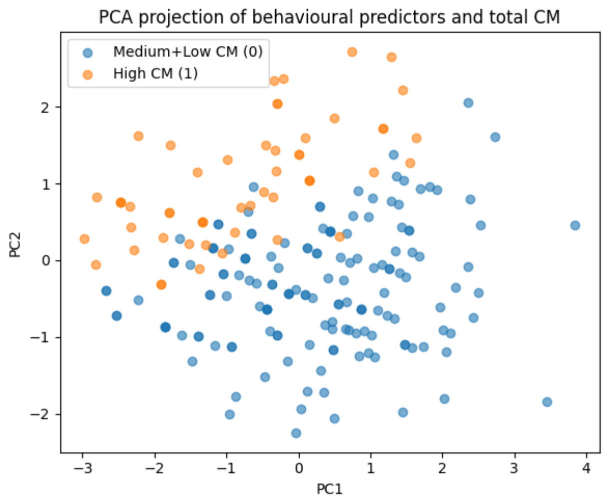

Figure 5: PCA projection of behavioural predictors and total CM.

Figure 5 displays the spatial distribution of observations in the space of the first two principal components constructed from the composite critical-thinking score and the three behavioral indicators of interaction with generative AI. Despite partial point overlap, students with high critical-thinking levels (class 1) are more frequently concentrated in a distinct region of the feature space than respondents with low/medium levels, confirming the presence of a structural pattern in the behavioral data.

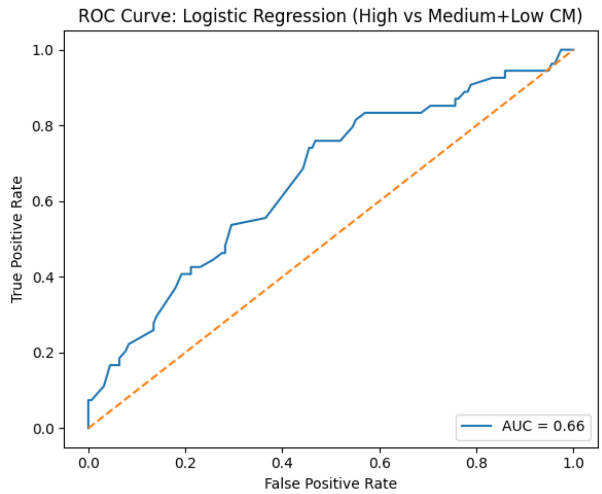

Figure 6: ROC Curve: Logistic Regression (High vs Medium+Low CM).

Figure 6 depicts the ROC curve for the logistic regression model classifying students into the high critical-thinking group. An AUC of 0.66 indicates a moderate yet statistically relevant ability of the model to distinguish between students with high versus lower levels of critical thinking using behavioral predictors alone.

Application of multinomial logistic regression showed that the probability of belonging to the High CM group increases significantly with greater frequency of AI-content verification (OR = 1.34) and with a higher willingness to seek consultation (OR = 1.62). By contrast, a higher degree of doubt regarding the accuracy of AI responses is associated with a lower likelihood of falling into the High CM group (OR = 0.77), which may reflect uncertainty/intellectual self-doubt that has not yet been translated into critical-thinking practice (Figure 7).

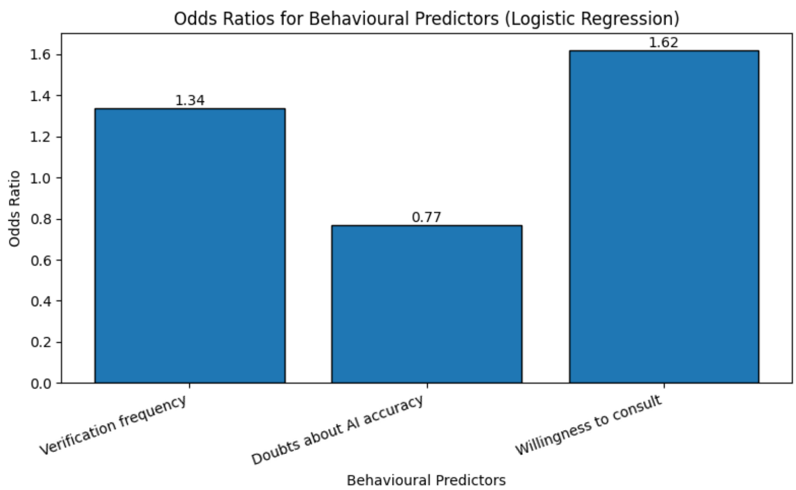

Figure 7: Odds ratios for behavioural predictors obtained from logistic regression model.

Thus, the logistic regression model clearly indicates that behavioral patterns of interaction with generative AI have predictive value for critical-thinking level. The strongest positive predictor is social verification – seeking consultation – which substantially increases the likelihood of belonging to the High CM group. The frequency of fact-checking likewise functions as a positive indicator. By contrast, doubt alone, without an accompanying verification process, does not cultivate a high level of critical judgment; rather, it is associated with a lower probability of achieving a high Watson–Glaser score.

5. Conclusions

The results suggest that the influence of generative AI on the development of critical thinking is non-linear and cannot be reduced to mere access to the technology. The decisive factor is the behavioral mode of interaction in which a student engages with AI as an information source. Our data show that the strongest links with high levels of critical thinking are verification-oriented actions – fact-checking, consulting alternative sources, and social verification through consultation. This implies that critical thinking is less an "internal trait" than a procedural activity manifested in concrete operations on information.

At the same time, the mere presence of doubt without subsequent steps to verify it does not serve as a marker of critical judgment. Such "untransformed uncertainty" reflects cognitive instability rather than cognitive complexity. This leads to an important conclusion: developing critical thinking in digital environments is not about cultivating more doubt, but about expanding the repertoire of control, verification, and analytical operations.

Accordingly, pedagogical interventions should not focus on restricting AI use or diminishing its role in learning, but on managing informational uncertainty. Building skills in verification, alternative interpretation, cross-checks, and social fact-checking can convert the "AI risk" into a resource for strengthening students' cognitive autonomy. In this sense, AI need not be a threat to critical thinking; it can function as a platform for its procedural reinforcement – provided that interaction with AI is reflective rather than passive.

Guided by the identified behavioral profiles, instructional efforts for class 0 (Non-High CM) students should prioritize transforming intuitive doubts into consistent verification actions. Each instance of uncertainty about an AI response should culminate in a brief fact-check using an independent source – ideally, two; the habit of "double-checking" should become a required step in completing assignments. It is also useful to practice formulating clarifying questions for the model, shifting from passive consumption of answers to the guided refinement of assumptions. Regular short discussions with peers or consultations with the instructor should be planned rather than ad hoc: once a week, select one AI response and run a micro–peer review. Such steps can convert scattered doubts into a structured thinking algorithm and gradually increase the likelihood of advancing to the High CM level.

For class 1 (High CM) students, the priority is to scale and stabilize already established practices. Formalize your verification procedures – write down fact-checking rules, source-selection criteria, and minimum evidentiary requirements – and apply them consistently across courses. Where feasible, turn individual checks into collaborative ones: organize brief reciprocal reviews in shared documents and record conclusions. For complex topics, raise the evidentiary bar – combine textbooks, peer-reviewed publications, and professional databases – and use AI as a counter-arguer explicitly tasked with probing weaknesses in your own reasoning. Institutionalizing these practices makes critical thinking not only an individual competence but also a collective norm within the learning community, thereby increasing error resilience and improving the quality of academic decisions.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] L. F. Santos, The role of critical thinking in science education, Journal of Education and Practice 8 (20) (2017) 159–173.

[2] B. Arisoy, B. Aybek, The effects of subject-based critical thinking education in mathematics on students' critical thinking skills and virtues, Eurasian Journal of Educational Research 92 (2021) 99–120. doi:10.14689/ejer.2021.92.6.

[3] T. Raj, P. Chauhan, R. Mehrotra, M. Sharma, Importance of critical thinking in the education, World Journal of English Language 12 (3) (2022) 126–133. doi:10.5430/wjel.v12n3p126.

[4] H. Pervaiz, K. Ali, S. Razzaq, M. Tariq, The impact of AI on critical thinking and writing skills in higher education, The Critical Review of Social Sciences Studies 3 (1) (2025) 3165–3176. doi:10.59075/79fkvy72.

[5] E. P. Ododo, U. B. Iniobong, A. I. Udoessien, I. U. Ukpe, O. D. James, Artificial intelligence in the classroom: perceived challenges to vocational education student retention and critical thinking in tertiary institutions, American Journal of Interdisciplinary Innovations and Research 6 (9) (2024) 30–39. doi:10.37547/tajiir/volume06issue09-05.

[6] K. Szmyd, E. Mitera, The impact of artificial intelligence on the development of critical thinking skills in students, European Research Studies Journal 27 (2) (2024) 1022–1039. doi:10.35808/ersj/3876.

[7] A. M. Vieriu, G. Petrea, The impact of artificial intelligence on students' academic development, Education Sciences 15 (3) (2025) 343. doi:10.3390/educsci15030343.

[8] M. Gerlich, AI tools in society: impacts on cognitive offloading and the future of critical thinking, Societies 15 (1) (2025) 6. doi:10.3390/soc15010006.

[9] H.-P. Lee, A. Sarkar, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, N. Wilson, The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25), Association for Computing Machinery, New York, NY, USA, 2025, Article 1121, 1–22. doi:10.1145/3706598.3713778.

[10] Q. Wang, A. H. Abdullah, Enhancing students' critical thinking through mathematics in higher education: a systemic review, SAGE Open 14 (3) (2024) 1–15. doi:10.1177/21582440241275651.