Babylonian Journal of Artificial Intelligence
Vol. **(2026)**, 2026, **pp**. 123–131
DOI: https://doi.org/10.58496/BJAI/2026/015 ISSN: 3006-5437
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Student Concerns and Perceived Negative Cognitive Effects of AI Tool Use in Academic Settings: A Text Mining Analysis

Rolaida L. Sonza[1], Argel S. Payumo[2,3], Marie Eugenie M. Otanes[4], Elmer M. Penecilla[5], Ivan G. Liwanag[4], Amado B. Martinez, Jr.[5], Jovita G. Rivera[5,6], John Paul P. Miranda[6,7,*]

[1] *Graduate School, Nueva Ecija University of Science and Technology, Nueva Ecija, Philippines*

[2] *National University, Philippines, Manila, Philippines*

[3] *School of Education, Arts and Sciences, National University, Philippines, Clark, Philippines*

[4] *College of Social Sciences and Philosophy, Pampanga State University, Pampanga, Philippines*

[5] *College of Education, Pampanga State University, Pampanga, Philippines*

[6] *Graduate School, Pampanga State University, Pampanga, Philippines*

[7] *College of Computing Studies, Pampanga State University, Pampanga, Philippines*

Article info

Article History
Received 20 Jun 2026
Revised 01 Nov 2026
Accepted 04 Nov 2026
Published 20 Nov 2026

Keywords
Artificial intelligence
Cognition
Text mining
AI dependency
Sentiment analysis

CC BY

ABSTRACT

The growing use of artificial intelligence (AI) tools in academic work has raised concerns about students' dependence on these tools and possible changes in their learning behaviors. This study examined perceived negative cognitive and behavioral changes reported by 197 college students in relation to their regular academic use of AI tools. Their open-ended responses were analyzed using word-frequency analysis, Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorization (NMF), and TextBlob sentiment analysis. Topic solutions from two to ten topics were compared based on coherence, consistency across repeated LDA runs, interpretability, and agreement between LDA and NMF. Two topics were retained: behavioral disengagement and AI dependency, and reduced independent cognitive and academic engagement. LDA assigned 130 participants (66%) to the first topic and 67 (34%) to the second, while NMF assigned 124 (62.9%) and 73 (37.1%), respectively. The two models assigned 177 participants (89.8%) to corresponding topics, with $\kappa = .779$. The mean TextBlob sentiment score was .090 (SD = .169). TextBlob scores did not differ significantly by sex, U = 4055, p = .081, and showed a small positive relationship with response length, $r_s = .266$, $p < .001$. The findings show that students commonly described dependence, reduced effort, and concerns about independent thinking.

1. INTRODUCTION

Artificial intelligence (AI) tools are increasingly used by students for academic tasks. These tools can generate explanations, summarize information, assist with writing, and provide immediate answers to academic questions [1], [2]. Their usefulness has encouraged students to integrate AI into everyday learning activities. However, AI use, particularly patterns of over-reliance, has also raised concerns about dependence, reduced effort, and reduced independent engagement with academic tasks [3], [4]. Students' perceptions of their academic engagement can provide useful information about how they experience changes in their learning practices [5]. AI dependency has emerged as a documented concern in higher education, with students reporting reliance on AI for tasks they might otherwise complete independently [6]. Studies of AI over-reliance have also identified concerns involving information seeking, problem solving, critical thinking, and cognitive offloading [7], [8]. Hasan [9] further argued that excessive AI assistance may reduce the effort involved in learning. Empirical studies have associated AI dependence or intensive AI-supported task completion with cognitive offloading, fatigue, reduced critical engagement, and lower independent effort [10]–[14].

Some students report that AI makes academic tasks easier while also describing greater reliance on automated assistance and less independent engagement with learning tasks [15]. Fan et al. [16] described a related pattern as metacognitive laziness,

*Corresponding author. Email: jppmiranda@pampangastateu.edu.ph

in which learners may transfer parts of their planning, monitoring, and evaluation to AI. Similar concerns have been raised about critical thinking, cognitive agency, and the possible replacement of student effort by automated assistance [11], [17]–[20]. These concerns do not mean that AI use necessarily causes cognitive decline, but they show that some students perceive changes in how they study and think. Written responses can provide useful information about these experiences. Structured surveys can measure predefined concepts, but open-ended questions allow students to describe concerns in their own words [21]. Text mining can then be used to identify recurring words and patterns across a large set of written responses [22]–[24]. Previous studies have used text mining to examine student feedback, engagement, perceptions, and other educational experiences [25]. Using more than one text-mining method can also provide a clearer view of whether similar patterns appear across different analytical approaches [26]–[28].

Recent reviews indicate that research on AI in higher education remains geographically concentrated, with developing-country contexts underrepresented in the available evidence [29], [30]. This limits the extent to which findings from other settings can be assumed to represent how Filipino college students use AI for academic work or describe its perceived effects. Differences in access, instructional practices, assessment requirements, and AI literacy may influence students' experiences. There is therefore a need to examine students' own accounts within this context. Such evidence can help teachers and institutions understand the concerns that students associate with regular academic AI use. This study examined students' self-reported perceptions of negative cognitive and behavioral effects associated with their regular academic use of AI tools. Specifically, it sought to answer the following questions: (1) What themes emerge from students' reports when analyzed using LDA and NMF? (2) What is the sentiment profile of the responses, and do sentiment scores differ significantly by sex? and (3) Is response length significantly related to students' sentiment scores?

2. MATERIALS AND METHODS

2.1 Research Design, Participants, Ethical Considerations, and Data Collection Procedure

This study used a descriptive cross-sectional design with text-mining methods [23]–[25], [31]. The analysis included 197 college students who reported negative cognitive or behavioral observations associated with their regular academic use of AI tools. The analysis was restricted to students who reported negative cognitive or behavioral observations because the study aimed to characterize these concerns rather than estimate their prevalence among all AI-using students. Participation was voluntary, informed consent was obtained from all participants, and ethics review was sought before data collection. Participants were informed of the purpose of the study and that their responses would be used for research purposes. The data were collected through open-ended survey questions, and the responses used for analysis did not contain personally identifying information.

The participants included 113 females (57.4%) and 84 males (42.6%). Their ages ranged from 18 to 28 years, with a mean age of 19.24 years (SD = 1.51). A total of 139 participants (70.6%) were aged 18-19 years, 41 (20.8%) were aged 20-21 years, and 17 (8.6%) were aged 22 years or older. The combined responses had a mean length of 59.04 words (SD = 43.14), with a median of 46 words and a range of 10-256 words. Table 1 presents the demographic profile and response length of the participants. The survey contained two open-ended questions. The first asked students to describe adverse effects that they associated with frequent academic AI use on their cognitive and behavioral patterns. The second asked them to describe changes that they had noticed in themselves while using AI tools. The two responses were combined for each participant for the word-frequency and topic-modeling analyses. For sentiment analysis, the responses were analyzed separately, and the two TextBlob scores were averaged for each participant.

TABLE I. DEMOGRAPHIC PROFILE AND RESPONSE LENGTH OF PARTICIPANTS

Variable	*f*	%	M (SD)
Sex			
Female	113	57.4	-
Male	84	42.6	-
Age Group			
18-19 years	139	70.6	-
20-21 years	41	20.8	-
22 years and above	17	8.6	-
Age (years)			19.24 (1.51)
Response Length (words)			
Question 1 and Question 2 Combined	197	-	59.04 (43.14)
Median			46 words
Range			10-256 words
Total	197	100	

2.2 Text Cleaning and Preparation

Text cleaning was conducted in Python through Jupyter Notebook. All responses were converted to lowercase, and punctuation and other non-letter characters were treated as word boundaries. Common English stopwords were removed together with AI-related terms such as AI, tool, tools, chatbot, and ChatGPT because these words occurred across many responses and provided little distinction between topics. Words that directly reflected the wording of the survey questions, including notice, change, affect, and effect, were also removed. The same cleaned text was used for word-frequency analysis, LDA, and NMF. Snowball stemming was used to group different forms of the same word. For example, study and studying were treated as the same word group. Depend, dependent, and dependence were also grouped, as were lazy and laziness and rely and relying. Fragments from contractions, such as *ve* from "*I've*," were removed because they did not carry useful meaning. Readable word forms are presented in the tables to make the results easier to interpret.

2.3 Word-Frequency and Topic-Modeling Procedures

Word-frequency analysis was first used to identify the words that appeared most often after text cleaning [25], [32]. LDA was then applied using word counts, while NMF used TF-IDF values to give more weight to words that were more informative within individual responses. LDA used unigram features, a minimum document frequency of three, a maximum vocabulary of 200 terms, 50 iterations, and a fixed random state for the final model. NMF used unigram and bigram features, a minimum document frequency of three, a maximum vocabulary of 300 terms, NNDSVDA initialization, and 500 maximum iterations. Each participant was assigned to the topic with the highest weight. Topic solutions from $k = 2$ to $k = 10$ were examined for both methods. C_V coherence was used to compare how closely the important words within each topic were related. LDA stability was assessed across five runs for each candidate solution from k = 2 to k = 10 using random seeds 0, 1, 2, 3, and 4. Topic similarity between runs was calculated using cosine similarity between the normalized topic-word distributions. Because topic numbering may change across runs, topics were matched using the Hungarian assignment algorithm to maximize the total similarity between corresponding topics. Stability for each candidate solution was calculated as the mean matched-topic cosine similarity across all 10 pairwise comparisons among the five runs. The final number of topics was selected based on coherence, consistency, agreement between LDA and NMF, clarity of the topics, and parsimony [26]–[28]. A sensitivity analysis was also performed for the word answer/answers. This word was retained in the main analysis because obtaining answers from AI may form part of students' reported reliance on these tools. LDA and NMF were then repeated after removing answer/answers. The topic assignments from both analyses were compared. This procedure tested whether the main topic structure depended heavily on one frequently occurring term.

2.4 Expert Review, Sentiment and Statistical Analysis

Two subject-matter experts independently validated the final two-topic solution. One specialized in educational technology and the other in educational psychology. Both experts reviewed the combined responses of all 197 participants and independently classified each response as either behavioral disengagement and AI dependency or reduced independent cognitive and academic engagement using the same topic definitions. The experts completed their ratings independently before any discussion. Inter-expert agreement was assessed using percentage agreement and Cohen's kappa [33].

TextBlob was used to estimate the positive or negative tone of the original responses [34]. The original text was retained because removing words and punctuation could alter the sentiment score. TextBlob assigned each response a score from -1.00 to +1.00, with scores greater than .05 classified as positive, scores from -.05 to .05 as neutral, and scores below -.05 as negative. The two question scores were averaged to obtain one sentiment score for each participant. To assess the validity of these classifications, a stratified random subset of 60 responses was selected, with 20 responses drawn from each TextBlob category. Two independent raters evaluated the contextual sentiment of each response as positive, neutral, or negative without access to the corresponding TextBlob classification, and agreement was assessed using percentage agreement and Cohen's kappa. The consensus manual classifications were then compared with the TextBlob classifications using the same measures. TextBlob scores were interpreted as lexical polarity indicators rather than validated contextual sentiment. Descriptive statistics were computed for participant characteristics, response length, and TextBlob scores. The Shapiro-Wilk test was used to assess normality. The Mann-Whitney U test was used as the primary test for sex differences, while the Welch t-test was used only as a sensitivity analysis.

3. RESULTS

3.1 Word-Frequency Analysis

Table 2 presents the 20 most frequently occurring word groups. Think/thinking had the highest frequency ($f = 123$), followed by answer/answers ($f = 111$), study/studying ($f = 110$), depend/dependent/dependence ($f = 108$), and lazy/laziness ($f = 102$). Rely/relying occurred 91 times. Other common terms included focus ($f = 69$), learn/learning ($f = 66$), motivation/motivated ($f = 56$), and feel/feeling ($f = 54$). Words related to thinking, dependence, study behavior, effort, and motivation were therefore prominent in the responses.

TABLE II. TOP 20 KEYWORDS BY FREQUENCY AND SEMANTIC CATEGORY

Rank	Keyword or related word forms	*f*	Category
1	think/thinking	123	Cognitive process
2	answer/answers	111	Task-oriented behavior
3	study/studying	110	Academic engagement
4	depend/dependent/dependence	108	Dependence
5	lazy/laziness	102	Behavioral disengagement
6	rely/relying	91	Dependence
7	make/making	84	Task-oriented behavior
8	focus	69	Attention
9	learn/learning	66	Academic engagement
10	instead	58	Behavioral substitution
11	time	58	Task management
12	motivation/motivated	56	Academic engagement
13	feel/feeling	54	Self-perception
14	work	48	Task-oriented behavior
15	understand/understanding	46	Cognitive process
16	critical/critically	46	Cognitive process
17	effort	38	Academic engagement
18	task	37	Task-oriented behavior
19	research/researching	37	Academic engagement
20	start/starting	35	Behavioral change

3.2 Selection of the Number of Topics

Table 3 presents the C_V coherence scores for topic solutions from k = 2 to k = 10. The two-topic LDA model had the highest coherence score (C_V = .620), followed by the three-topic model (C_V = .610) and the six-topic model (C_V = .590). NMF had its highest coherence at eight topics (C_V = .639), while the two-topic solution produced C_V = .592. However, the eight-topic NMF solution produced narrower and less distinct topic groupings. The two-topic structure was retained because it showed the strongest LDA coherence, high stability across repeated LDA runs, comparable patterns across both models, and clearer topic separation. The two-topic LDA model showed the highest stability across the five repeated runs, with a mean matched-topic cosine similarity of 1.000. Stability was .902 for three topics, .821 for four topics, .763 for five topics, .754 for six topics, .732 for seven topics, .716 for eight topics, .716 for nine topics, and .686 for ten topics. These results provided additional support for retaining the two-topic solution in the final LDA and NMF analyses.

TABLE III. CV COHERENCE AND LDA STABILITY ACROSS CANDIDATE TOPIC SOLUTIONS

Number of topics	LDA C_V	NMF C_V	LDA stability
2	0.620	0.592	1.000
3	0.610	0.547	0.902
4	0.575	0.624	0.821
5	0.579	0.587	0.763
6	0.590	0.610	0.754
7	0.550	0.584	0.732
8	0.529	0.639	0.716
9	0.552	0.579	0.716
10	0.531	0.590	0.686

3.3 LDA Topic Results

Table 4 presents the two topics identified through LDA. Behavioral disengagement and AI dependency included 130 participants (66%). Reduced independent cognitive and academic engagement included 67 participants (34%). The first topic contained words such as lazy, depend, rely, work, study, answer, and activity. The second contained words such as think, study, focus, learn, critical, motivation, and understand.

TABLE IV. LDA-DERIVED TOPICS

Rank	Topic label	Representative keywords	*f*	%
1	Behavioral disengagement and AI dependency	lazy, depend, think, work, feel, rely, make, study, answer, longer, know, activity	130	66
2	Reduced independent cognitive and academic engagement	think, answer, study, rely, focus, learn, critical, make, instead, motivation, understand, depend	67	34

3.4 NMF Topic Results

Table 5 presents the results of the NMF analysis. Behavioral disengagement and AI dependency included 124 participants (62.9%), while reduced independent cognitive and academic engagement included 73 participants (37.1%). The first topic emphasized lazy, depend, work, assignment, rely, and activity. The second emphasized answer, think, study, critical, motivation, focus, understand, learn, and research. The distributions were similar to those produced by LDA. The two models assigned 177 of the 197 participants (89.8%) to corresponding topics. Cohen's kappa was .779, while the adjusted Rand index was .631. After answer/answers was removed, 196 of the 197 LDA assignments (99.5%) and 193 of the 197 NMF assignments (98%) remained in the corresponding topic. In this sensitivity analysis, LDA assigned 131 participants to the first topic and 66 to the second. NMF assigned 120 participants to the first topic and 77 to the second. The two experts agreed on the topic classification of 180 of the 197 responses, corresponding to 91.4% agreement. Cohen's κ was .818, indicating almost perfect inter-expert agreement. Expert 1 classified 118 responses as Topic 1 and 79 as Topic 2, while Expert 2 classified 123 as Topic 1 and 74 as Topic 2. The experts differed on 17 responses.

TABLE V. NMF-DERIVED TOPICS

Rank	Topic label	Representative keywords	*f*	%
1	Behavioral disengagement and AI dependency	lazy, depend, feel, think, longer, lazy thinking, work, assignment, know, make, rely, activity	124	62.9
2	Reduced independent cognitive and academic engagement	answer, think, study, critical, instead, motivation, focus, understand, rely, learn, research, quick	73	37.1

3.5 Sentiment Analysis Results

Table 6 presents the TextBlob sentiment classifications. The mean score for Question 1 was .068, while the mean for Question 2 was .111, producing a composite mean of .090 (SD = .169). TextBlob classified 109 responses (55.3%) as positive, 47 (23.9%) as neutral, and 41 (20.8%) as negative. Manual validation produced a different pattern, as the two independent raters agreed on all 60 responses in the validation subset, yielding 100% agreement and Cohen's κ = 1. They classified 54 responses (90%) as negative and 6 (10%) as neutral, with no positive classifications. The consensus manual ratings agreed with TextBlob in 22 of 60 cases (36.7%), with Cohen's κ = .050, indicating slight agreement. Among the TextBlob-positive responses, 18 were manually classified as negative and 2 as neutral; among the TextBlob-neutral responses, 17 were classified as negative and 3 as neutral; and among the TextBlob-negative responses, 19 were classified as negative and 1 as neutral.

TABLE VI. SENTIMENT POLARITY DISTRIBUTION OF STUDENT RESPONSES

Sentiment category	Score	*f*	%
Positive	> .05	109	55.3
Neutral	-.05 to .05	47	23.9
Negative	< -.05	41	20.8
Total	-	197	100

3.6 Statistical Comparisons

Table 7 presents the statistical comparisons based on the TextBlob scores. The Shapiro-Wilk test showed no significant departure from normality, $W = .989$, $p = .156$. Female participants had a median TextBlob score of .106 (IQR = .235), while male participants had a median of .058 (IQR = .249). The primary Mann-Whitney U test showed no significant difference, $U = 4055$, $p = .081$, with a small rank-biserial effect size of $|r_{rb}| = .146$. The Welch t-test, conducted as a sensitivity analysis, produced a similar result, $t(176.12) = 1.80$, p = .074, Hedges' g = .26. Sex was also not significantly associated with dominant LDA topic membership, $\chi^2(1) = 1.93$, $p = .165$, Cramér's $V = .099$. Among female participants, 70 were assigned to behavioral disengagement and AI dependency and 43 to reduced independent cognitive and academic engagement. Among male participants, 60 were assigned to the first topic and 24 to the second. Response length showed a small positive relationship with the TextBlob score, $r_s = .266$, p < .001.

TABLE VII. SUMMARY OF STATISTICAL COMPARISONS

Analysis	Statistic	*p*	Effect size	Result
Sex and dominant LDA topic	$\chi^2(1) = 1.93$	.165	$V = .099$	Not significant
TextBlob score by sex	$U = 4055$	.081	$\lvert r_{rb}\rvert = .146$	Not significant
Welch sensitivity test	$t(176.12) = 1.80$	.074	$g = .26$	Not significant
Response length and TextBlob score	$r_s = .266$	< .001	$r_s = .266$	Significant

4. DISCUSSION

The topic-modeling results identified two broad patterns in students' accounts of perceived negative effects associated with academic AI use. The first involved behavioral disengagement and AI dependency, while the second involved reduced independent cognitive and academic engagement. Behavioral disengagement and AI dependency was the larger topic in both LDA and NMF and included references to laziness, reliance, reduced work, assignments, and dependence on AI-generated assistance. These findings are consistent with studies that associate excessive AI reliance with cognitive offloading, reduced self-regulation, and lower independent effort [6], [11]–[14], [16]. Fan et al. [16] described a related pattern as metacognitive laziness, in which learners transfer parts of their planning and monitoring to AI. The second topic focused more on thinking, focus, critical thinking, learning, motivation, and understanding, which are also reflected in previous studies on AI-related cognitive offloading and critical-thinking processes [7], [10], [12], [35]. These findings reflect students' self-reported perceptions and should not be interpreted as evidence of objectively measured cognitive change or impairment. The similarity between the two models further supports the identified patterns, as LDA and NMF assigned 89.8% of participants to corresponding topics. The sensitivity analysis also showed that removing answer/answers produced very little change in participant assignments, indicating that the two-topic structure was not dependent on this single frequent term.

The sentiment findings differed substantially between the automated and manual classifications. Although TextBlob classified 55.3% of the responses as positive, the manual validation showed that most responses in the validation subset were negative when their overall contextual meaning was considered. The two raters showed complete agreement, while agreement between the consensus manual classifications and TextBlob was only 36.7%, with $\kappa = .050$. The low agreement observed in the present study shows that the TextBlob classifications did not consistently capture the contextual meaning of these responses. Statements describing dependence, reduced motivation, difficulty concentrating, or reduced independent thinking may contain words with positive or neutral lexical values even when the overall experience is negative. The manual validation supports a cautious interpretation of the TextBlob results. The human ratings provide the more appropriate contextual assessment, while the TextBlob scores primarily reflect lexical polarity rather than validated contextual sentiment. No significant differences were found between female and male participants in the TextBlob scores or dominant topic membership. Response length showed only a small positive relationship with the TextBlob score.

The findings have practical implications for how AI use is managed in higher education. Guidance should focus not only on whether students are allowed to use AI but also on which parts of academic work should remain under the student's responsibility. Teachers can require students to explain their reasoning, compare their own answers with AI-generated responses, justify revisions, or complete selected stages of a task independently before using AI support. These activities can preserve opportunities for critical thinking and self-regulation while allowing students to benefit from appropriate AI assistance [16], [19], [20]. AI literacy programs can also help students recognize when convenience begins to replace independent effort or sustained engagement with a learning task. Assessment design can support the same goal by requiring explanation, verification, synthesis, reflection, or personal justification rather than outputs that can be produced with little student involvement. Teachers may also ask students to identify which parts of an academic task were completed with AI and which were completed independently. Institutional policies can support these practices by distinguishing productive AI assistance from patterns of use that may replace important learning processes [2], [4]. This approach recognizes that students may value the usefulness of AI while also expressing concerns about dependence or reduced learning.

The sample included only students who reported negative cognitive or behavioral observations, so the findings are not representative of all college students who use AI and cannot be used to estimate the prevalence of these concerns. Its design and reliance on self-reported experiences also prevent conclusions about whether AI use caused the changes described by participants [31]. The study was conducted within one national and educational context, and the findings may differ in institutions with other student populations, levels of AI access, academic practices, or institutional policies. Topic-modeling results can also vary according to text cleaning, vocabulary selection, model settings, and the number of topics retained. The use of alternative topic solutions, repeated LDA runs, two topic-modeling methods, and a sensitivity analysis reduced some of these concerns but could not remove all analytical choices. TextBlob also has limitations because it relies largely on word-level sentiment and cannot fully capture context, implied criticism, or mixed evaluations [34]. The wording of the survey questions may likewise influence some of the words students used even after directly related terms were removed. Future studies can address these limitations by including students with positive, neutral, and negative AI experiences and by combining self-reported accounts with longitudinal or performance-based measures of learning and cognitive engagement.

5. CONCLUSION

This study examined students' reported negative cognitive and behavioral experiences associated with regular academic AI use. Text mining identified two main patterns: behavioral disengagement and AI dependency, and reduced independent cognitive and academic engagement. LDA and NMF produced similar topic distributions, and most participants were assigned to corresponding topics by both methods. The topic structure also remained stable when answer/answers was removed from the analysis. Although TextBlob produced a mildly positive average sentiment score, manual validation

showed that its classifications had only slight agreement with contextual human judgments, indicating that the automated scores mainly reflected surface lexical tone rather than the overall meaning of the responses. TextBlob scores and topic membership did not differ significantly by sex, while response length showed only a small relationship with the TextBlob score. The findings describe students' perceived changes in effort, dependence, and independent thinking rather than objectively measured cognitive change. Future studies should include students with positive, neutral, and negative AI experiences and use longitudinal, qualitative, or performance-based measures to examine whether these perceived changes are associated with measurable differences in learning and cognitive engagement.

Acknowledgements

The authors would like to express their sincere gratitude to all the participants who generously took part in this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Authorship Contribution Statement

Rolaida L. Sonza: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision. **Argel S. Payumo:** Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Marie Eugenie M. Otanes:** Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Elmer M. Penecilla:** Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Ivan G. Liwanag:** Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Amado B. Martinez, Jr.:** Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Jovita G. Rivera:** Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **John Paul P. Miranda:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Supervision. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval

The study obtained an official Certification of Exemption from Research Ethics Review from the Graduate School Ethics Review Panel of Nueva Ecija University of Science and Technology on February 4, 2026. Upon review of the study protocol, the Panel determined that the study qualified for exemption under Provision 47, page 48, of the National Ethical Guidelines for Health and Health-Related Research (2022). Informed consent was obtained from all participants electronically through an online survey form prior to their participation in the study.

Data Availability Statement

The data supporting the findings of this study are available from corresponding author upon reasonable request.

References

[1] D. A. Schmidt, B. Alboloushi, A. Thomas, and R. Magalhaes, "Integrating artificial intelligence in higher education: perceptions, challenges, and strategies for academic innovation," *Comput. Educ. Open*, vol. 9, p. 100274, 2025, doi: 10.1016/j.caeo.2025.100274.

[2] B. Klimova and M. Pikhart, "Exploring the effects of artificial intelligence on student and academic well-being in higher education: a mini-review," *Front. Psychol.*, vol. 16, p. 1498132, 2025, doi: 10.3389/fpsyg.2025.1498132.

[3] D. Lee *et al.*, "The impact of generative AI on higher education learning and teaching: A study of educators' perspectives," *Comput. Educ. Artif. Intell.*, vol. 6, p. 100221, 2024, doi: 10.1016/j.caeai.2024.100221.

[4] S. Abubakar, A. Jeilani, and M. Yusuf, "The role of over-reliance on AI in the negative consequences of student learning: The moderating effects of ethical concerns and institutional policies," *Cogent Educ.*, vol. 12, no. 1, p. 2591503, Dec. 2025, doi: 10.1080/2331186X.2025.2591503.

[5] C. M. Amerstorfer and C. Freiin von Münster-Kistner, "Student Perceptions of Academic Engagement and Student-Teacher Relationships in Problem-Based Learning.," *Front. Psychol.*, vol. 12, p. 713057, 2021, doi: 10.3389/fpsyg.2021.713057.

[6] S. Zhang, X. Zhao, T. Zhou, and J. H. Kim, "Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior," *Int. J. Educ. Technol. High. Educ.*, vol. 21, no. 1, p. 34, 2024, doi: 10.1186/s41239-024-00467-0.

[7] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review," *Smart Learn. Environ.*, vol. 11, no. 1, p. 28, 2024, doi: 10.1186/s40561-024-00316-7.

[8] F. Noorbehbahani and K. Oyibo, "AI-overdependence and human cognitive decline: Hazards, evidence, and mitigation strategies," *Comput. Hum. Behav. Reports*, vol. 22, p. 101102, 2026, doi: 10.1016/j.chbr.2026.101102.

[9] M. K. Hasan, "How AI quietly undermines the joy and effort of learning: a call for rebalancing education in the digital age.," *Ann. Med. Surg.*, vol. 87, no. 8, pp. 4693–4694, Aug. 2025, doi: 10.1097/MS9.0000000000003456.

[10] J. Tian and R. Zhang, "Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy," *Acta Psychol. (Amst).*, vol. 260, p. 105725, 2025, doi: 10.1016/j.actpsy.2025.105725.

[11] B. Jose, J. Cherian, A. M. Verghis, S. M. Varghise, M. S, and S. Joseph, "The cognitive paradox of AI in education: between enhancement and erosion," *Front. Psychol.*, vol. 16, 2025, doi: 10.3389/fpsyg.2025.1550621.

[12] M. Gerlich, "AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking," *Societies*, vol. 15, no. 1, p. 6, 2025, doi: 10.3390/soc15010006.

[13] Y. Jiang and H. Chen, "AI-based cognitive offloading by students in higher education in China," *Comput. Hum. Behav. Artif. Humans*, vol. 8, p. 100327, 2026, doi: 10.1016/j.chbah.2026.100327.

[14] K. Misiejuk, S. López-Pernas, R. Kaliisa, and M. Saqr, "Cognitive offloading in student–AI collaboration: A longitudinal analysis of prompting strategies," *Comput. Hum. Behav. Reports*, vol. 22, p. 101130, 2026, doi: 10.1016/j.chbr.2026.101130.

[15] G. Krsmanović and F. P. Deek, "Self-Reported Cognitive Effects of AI Applications on Learning," *J. High. Educ. Theory Pract.*, vol. 25, no. 1, pp. 55–62, 2025, doi: 10.33423/jhetp.v25i1.7563.

[16] Y. Fan *et al.*, "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance," *Br. J. Educ. Technol.*, vol. 56, no. 2, pp. 489–530, 2024, doi: 10.1111/bjet.13544.

[17] X. Peng, K. Sun, X. Shan, and J. Zhang, "Empowerment or dependency? A systematic review of the impacts of intelligent assessment and generative AI on learners' self-beliefs and cognitive agency in music education," *Front. Psychol.*, vol. 17, p. 1776445, 2026, doi: 10.3389/fpsyg.2026.1776445.

[18] L. Konradt, D. N. Boote, and M. Taub, "Making AI annoying on purpose: when helpful tools don't always help," *Comput. Educ. Open*, vol. 10, p. 100343, 2026, doi: 10.1016/j.caeo.2026.100343.

[19] M. Vendrell and S.-K. Johnston, "Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education," *Comput. Educ. Artif. Intell.*, vol. 10, p. 100572, 2026, doi: 10.1016/j.caeai.2026.100572.

[20] L. Ding, R. O'Berry, H. Tallent, S. G. Chhetri G C, and A. Gapud, "Learning with large language models: beyond prompt engineering," *Educ. Inf. Technol.*, vol. 31, no. 9, pp. 2877–2901, 2026, doi: 10.1007/s10639-026-13924-2.

[21] C. Hamshire *et al.*, "The potential of student narratives to enhance quality in higher education," *Qual. High. Educ.*, vol. 23, no. 1, pp. 50–64, Jan. 2017, doi: 10.1080/13538322.2017.1294407.

[22] N. Zipser, D. Kurochkin, K. W. Yu, and L. A. Mincieli, "Exploring Relationships Between Qualitative Student Evaluation Comments and Quantitative Instructor Ratings: A Structural Topic Modeling Framework," *Educ. Sci.*, vol. 15, no. 8, p. 1011, 2025, doi: 10.3390/educsci15081011.

[23] Ardyn Nordstrom, Aditya Maheshwari, Rheann Quenneville, and Grace Shen-Tu, "Combining Text Mining and Manual Thematic Analysis to Understand Participant Experiences With Surveys," *Int. J. Qual. Methods*, vol. 24, p. 16094069251378860, Jul. 2025, doi: 10.1177/16094069251378860.

[24] A. O'Mara-Eves, J. Thomas, J. McNaught, M. Miwa, and S. Ananiadou, "Using text mining for study identification in systematic reviews: a systematic review of current approaches.," *Syst. Rev.*, vol. 4, no. 1, p. 5, Jan. 2015, doi: 10.1186/2046-4053-4-5.

[25] R. Ferreira-Mello, M. André, A. Pinheiro, E. Costa, and C. Romero, "Text mining in education," *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 6, p. e1332, Nov. 2019, doi: 10.1002/widm.1332.

[26] M. C. Schlunegger, M. Zumstein-Shaha, and R. Palm, "Methodologic and Data-Analysis Triangulation in Case Studies: A Scoping Review.," *West. J. Nurs. Res.*, vol. 46, no. 8, pp. 611–622, Aug. 2024, doi: 10.1177/01939459241263011.

[27] M. Couto, J. Parapar, and D. E. Losada, "Exploiting topic analysis models to explore psychological dimensions in social media data.," *Sci. Rep.*, vol. 16, no. 1, p. 6047, Jan. 2026, doi: 10.1038/s41598-026-36339-y.

[28] A. Nugumanova, A. Alzhanov, A. Mansurova, K. Rakhymbek, and Y. Baiburin, "Semantic Non-Negative Matrix Factorization for Term Extraction," *Big Data and Cognitive Computing*, vol. 8, no. 7. p. 72, 2024. doi: 10.3390/bdcc8070072.

[29] J. Liang, J. M. Stephens, and G. T. L. Brown, “A systematic review of the early impact of artificial intelligence on higher education curriculum, instruction, and assessment,” *Front. Educ.*, vol. 10, p. 1522841, 2025, doi: 10.3389/feduc.2025.1522841.

[30] D. Y. Long, S. Wang, S. Md Rashid, and X. T. Lu, “Artificial intelligence in higher education: a systematic review of its impact on student engagement and the mediating role of teaching methods,” *Front. Educ.*, vol. 10, p. 1648661, 2026, doi: 10.3389/feduc.2025.1648661.

[31] X. Wang and Z. Cheng, “Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations.,” *Chest*, vol. 158, no. 1S, pp. S65–S71, Jul. 2020, doi: 10.1016/j.chest.2020.03.012.

[32] S. Kang and S. Kim, “Lessons Learned from Topic Modeling Analysis of COVID-19 News to Enrich Statistics Education in Korea,” *Sustainability*, vol. 14, no. 6, p. 3240, 2022, doi: 10.3390/su14063240.

[33] J. R. Landis and G. G. Koch, “The Measurement of Observer Agreement for Categorical Data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, May 1977, doi: 10.2307/2529310.

[34] S. In and S. Chanchamnan, “Comparison of VADER and TextBlob labeling for sentiment analysis using machine learning and deep learning models: A study on generative AI user experience.,” *Acta Psychol. (Amst).*, vol. 263, p. 106268, Mar. 2026, doi: 10.1016/j.actpsy.2026.106268.

[35] P. Ninghardjanti, M. C. Umam, A. Subarno, W. Winarno, N. R. Langgi, and J. Widodo, “Evaluating the impact of AI on the critical thinking skills among the higher education students by combining the TAM model and critical thinking theory,” *Front. Educ.*, vol. 10, p. 1719625, 2025, doi: 10.3389/feduc.2025.1719625.