NMIRJ
Nrityanjali Multidisciplinary International Research Journal
Volume 2, Issue 2, April -June 2026
E-ISSN: 3108-1452
Website: https://nrityanjaliresearchjournal.com/home/ Email: info@nrityanjaliresearchjournal.com

Artificial Intelligence in the Education Sector: A Threat to Human Cognitive Abilities or a Better Tool for Learning

Ruchi Charak
Department of BCA, School of Engineering and Computer Technology
Cluster University of Jammu.
Emailid- charakruchi1104@gmail.com
Mobile-7051930903

https://doi.org/10.5281/zenodo.22057391

Abstract

The diffusion of generative artificial intelligence (AI) into classrooms has produced two competing narratives: that AI is the most powerful personalisation engine education has ever had, and that reliance on machine-generated answers hollows out the capacities education exists to build. This study tested both in a single framework. A cross-sectional survey of 420 higher education students measured AI usage intensity, cognitive offloading, critical thinking disposition, metacognitive self-regulation, AI literacy, perceived learning benefit and academic performance (α = .851 to .910), analysed through correlation, t tests, ANOVA, chi-square, multiple regression and bootstrapped mediation. A clear paradox emerged. Perceived benefit was high ($M = 3.89$, $SD = 0.65$), yet AI usage correlated negatively with critical thinking ($r = -.734$, $p < .001$) and positively with cognitive offloading ($r = .685$, $p < .001$). Heavy users scored far lower on critical thinking than light users, $F(2, 417) = 172.01$, $p < .001$, $\eta^2 = .452$. Regression explained 73.8% of variance in critical thinking, with metacognitive self-regulation the strongest positive predictor ($\beta = .408$) and AI literacy non-significant. Offloading carried 39.6% of the total effect (bootstrap 95% CI [-.385, -.248]).

Keywords: artificial intelligence in education; cognitive offloading; critical thinking; generative AI; metacognitive self-regulation; higher education

Introduction

Education has always absorbed its tools, and writing, print, the calculator and the search engine each arrived with anxieties about what learners would stop being able to do. Generative artificial intelligence differs in one respect: earlier tools stored or transported knowledge, whereas large language models produce the intellectual artefact itself. A calculator returns a number the student must still interpret; a generative model returns the argument the student was supposed to construct, moving the technology from the periphery of the learning task to its centre.

Adoption has been extraordinarily rapid. Within two years of public release, conversational AI had become routine in student workflow across most disciplines and regions (Maslej et al., 2025), while institutional response lagged behind student behaviour: UNESCO issued the first global guidance for generative AI in education only after mass adoption was under way, noting that the absence of national frameworks left learners and their data unprotected (Miao & Holmes, 2023).

The evidence points in two directions at once. Meta-analyses report substantial positive effects on learning outcomes: $g+ = 0.712$ for academic performance across 69 experimental studies (Deng et al., 2025) and $g = 0.857$ for learning achievement across 49 studies (Liu et al., 2025). Against this, a growing body of work reports that frequent AI use is associated with reduced critical thinking, weaker neural engagement and a tendency to delegate evaluative judgement to the machine (Gerlich, 2025; Kosmyna et al., 2025; Lee et al., 2025). Both bodies are methodologically respectable, so the task is not to choose between them but to identify the conditions under which each holds.

For curriculum designers the study identifies metacognitive self-regulation rather than restriction as the operative lever; for teachers it suggests assessment should make offloaded steps gradeable; for policy bodies it supplies evidence from a student population that can be set against the largely Western samples dominating the literature. It covers higher education students during 2025-26 and is confined to text-based generative AI assistants; critical thinking is measured as a self-reported disposition and academic performance as self-reported cumulative grade point average.

AI Usage Intensity (AIU) denotes the self-reported frequency, breadth and depth of use of generative AI for academic tasks. Cognitive Offloading (COF) denotes the extent to which recall, computation, synthesis and evaluation are transferred to an external AI system instead of being performed internally (Gerlich, 2025; Zhai et al., 2024). Critical Thinking Disposition (CTD) denotes the habitual inclination to question sources, weigh alternatives and suspend judgement pending verification. Metacognitive Self-Regulation (MSR) denotes the capacity to plan, monitor and evaluate one's own cognition, including when to withhold a tool.

Review of Literature

Theoretical Framework

Four theoretical positions frame the enquiry. Cognitive load theory holds that working memory is limited and instruction should reduce extraneous load so germane load can be devoted to schema construction; on this reading an AI assistant that removes mechanical drudgery is beneficial, and Deng et al. (2025) offer partial support in reporting reduced mental effort (g+ = -0.675) alongside performance gains. Distributed cognition instead treats the tool as part of the cognitive system, so the question is whether the human-tool assembly performs well; the difficulty, as Lee et al. (2025) show, is that the assembly is only as reliable as the human's willingness to verify machine output, and that willingness declines as confidence in the tool rises.

Research on desirable difficulties supplies the strongest basis for concern: retrieval effort and productive struggle are the conditions under which durable learning occurs, and a tool that eliminates the struggle eliminates the mechanism. Fan et al. (2025) provide the sharpest test, showing ChatGPT-supported students produced the largest essay-score improvement yet gained no advantage in knowledge acquisition or transfer — a dissociation they label metacognitive laziness. Self-regulated learning theory supplies the moderating construct: learners who plan, monitor and evaluate their own processing use AI instrumentally, whereas those without such regulation are used by it. Shi et al. (2025) found self-regulated learning a stronger predictor of writing performance (β = .237) than AI literacy (β = .153).

Artificial Intelligence as an Enabler of Learning

The evidence for benefit is substantial and converging. Ma and Zhong (2025) synthesised 34 studies and obtained a combined effect of 0.68 on learning outcomes — cognitive (g = 0.795), competency (g = 0.711) and affective (g = 0.507) — so generative AI does most for knowledge acquisition and least for attitudes. Nie et al. (2025), in a three-level meta-analysis of 68 effect sizes from 2,347 participants, found a large positive effect on higher-order thinking itself (g = 0.851, 95% CI [0.452, 1.250]).

Three qualifications temper this picture. Liu et al. (2025) found the motivational advantage diminished as intervention duration increased, consistent with a novelty effect; Deng et al. (2025) found no significant effect on academic self-efficacy (g+ = 0.441, 95% CI [-0.141, 1.023]); and Létourneau et al. (2025) found that when AI-driven tutoring systems were compared against non-intelligent tutoring systems rather than traditional teaching, the advantage largely disappeared, implying much of the measured benefit is attributable to tutoring structure rather than to artificial intelligence.

Artificial Intelligence as a Threat to Cognitive Capacity

The countervailing literature is younger but methodologically diverse. Gerlich (2025) surveyed 666 participants and conducted 50 interviews, reporting a strong negative correlation between frequent AI use and critical thinking (r = -.68), a positive correlation between use and cognitive offloading (r = .72), and a negative association between offloading and critical thinking (r = -.75); offloading mediated the relationship and higher educational attainment buffered the effect.

Neurophysiological evidence has since been added. Kosmyna et al. (2025) recorded electroencephalographic activity from 54 participants writing essays under three conditions and found neural connectivity scaled inversely with external support: brain-only writers showed the strongest coupling and language-model users the weakest. The language-model group reported lower ownership of their essays, were frequently unable to quote work submitted minutes earlier, and when switched to unaided writing showed reduced alpha and beta connectivity — accumulated cognitive debt. Lee et al. (2025), surveying 319 knowledge workers supplying 936 real instances of AI use, found higher confidence in the AI predicted less critical thinking effort while higher confidence in one's own expertise predicted more.

Over-reliance, Academic Integrity and Metacognition

Zhai et al. (2024), reviewing 14 studies of over-reliance on AI dialogue systems, concluded that habitual reliance degrades decision-making, critical thinking and analytical reasoning as users substitute rapid heuristic acceptance for deliberate evaluation, with a 75% risk rating attached to diminished critical thinking. Niloy et al. (2024), modelling 594 South Asian students, found motivations for chatbot use significantly associated with declining academic integrity, with rote-oriented pedagogy the strongest driver of substitution — a finding that relocates part of the problem from student character to assessment design.

Evidence from Asian and Indian Higher Education

Ma et al. (2026) surveyed 443 Chinese university students and modelled two opposing indirect pathways from AI use to academic performance — a positive path through academic self-efficacy and a negative path through thinking offloading — a double-edged structure that anticipates the framework tested here. Singh et al. (2026), working with 60 students and 20 teachers across Indian universities, found the most salient barrier for both groups was the absence of institutional guidance. Nasr et al. (2025) found students agreed ChatGPT supported the triggering ($M = 3.60$), exploration ($M = 3.70$) and integration ($M = 3.60$) phases of inquiry but were neutral about resolution ($M = 3.30$).

Policy Frameworks and Research Gap

Institutional guidance has begun to converge on capability building rather than prohibition. UNESCO's competency framework for students organises the required capabilities into a human-centred mindset, AI ethics, AI techniques and applications, and AI system design (Miao et al., 2024). Three gaps nevertheless remain: the benefit literature measures task performance while the harm literature measures cognitive process, and the two are rarely assessed in the same sample; the mediating mechanism proposed by Gerlich (2025) and Ma et al. (2026) has not been replicated in an Indian higher education population; and almost no study has tested metacognitive self-regulation and AI literacy simultaneously as competing protective factors. The present study addresses all three.

Research Methodology

Research Design, Population and Sampling

A quantitative, non-experimental, cross-sectional survey design with a descriptive and causal-comparative orientation was adopted, because the research questions concern naturally occurring variation rather than a manipulated intervention, and because the mediation hypothesis requires simultaneous measurement of predictor, mediator and outcome in one sample. The target population comprised undergraduate, postgraduate and research students during 2025-26, and stratified purposive sampling with discipline as the stratifying variable ensured representation across four broad streams. Of 486 questionnaires distributed, 431 were returned and 420 retained after removal of incomplete and patterned responses, an effective response rate of 86.4% that exceeds the minimum of 384 required for 95% confidence at a 5% margin of error.

Table 1
Demographic Profile of Respondents ($N = 420$)

Variable	Category	Frequency	Percentage
Gender	Male	211	50.2
	Female	209	49.8
Age group (years)	18–20	169	40.2
	21–23	141	33.6
	24–26	71	16.9
	27 and above	39	9.3
Level of study	Undergraduate	209	49.8
	Postgraduate	157	37.4
	Research scholar	54	12.9
Discipline	Science & Engineering	145	34.5
	Commerce & Management	117	27.9
	Humanities & Social Sciences	90	21.4
	Professional streams	68	16.2

Note. Percentages are rounded to one decimal place.

Research Instrument, Validity and Reliability

The questionnaire had three sections: Section A recorded demographics, Section B captured frequency, purpose and duration of AI use, and Section C contained 37 Likert-type statements measuring the six latent constructs on a five-point agreement scale anchored at 1 (strongly disagree) and 5 (strongly agree). Items were adapted from instruments reported by Gerlich (2025), Shi et al. (2025) and Ma et al. (2026), and negatively worded items were reverse-scored before analysis.

Content validity was established by five subject experts, with a content validity ratio above .62 for all retained items, and a pilot on 45 respondents excluded from the final sample led to rewording four ambiguous items. Exploratory factor analysis supported construct validity: the Kaiser-Meyer-Olkin measure was .887, Bartlett's test was significant, $\chi^2(666) = 8{,}142.36$, $p < .001$, and six factors with eigenvalues above unity explained 68.4% of total variance, all items loading above .50 on their intended factor (Table 2).

Table 2
Construct Composition and Reliability Statistics

Construct	Code	Items	Cronbach's α	Item–total r
AI Usage Intensity	AIU	6	.860	.64–.66
Cognitive Offloading	COF	6	.870	.65–.69
Critical Thinking Disposition	CTD	8	.910	.69–.74
Metacognitive Self-Regulation	MSR	6	.877	.67–.70
AI Literacy	AIL	5	.860	.65–.71
Perceived Learning Benefit	PLB	6	.851	.60–.69
Full instrument	—	37	.893	—

Note. All α values exceed the .70 criterion and all corrected item–total correlations exceed the .30 retention threshold.

Data Collection, Statistical Techniques and Ethics

Data were collected over eight weeks using a dual-mode strategy: a digital form circulated through institutional mailing lists, and printed copies administered in classrooms to reduce self-selection bias. Analysis used descriptive statistics, Pearson correlation, independent-samples t tests with Levene's correction, one-way ANOVA with Tukey post hoc comparison, chi-square tests of independence, multiple regression, and mediation analysis using the causal steps procedure with a bias-corrected bootstrap of 5,000 resamples. Assumptions of normality, linearity, homoscedasticity and absence of multicollinearity were verified before each parametric test, with significance at $\alpha = .05$. Participation was voluntary and anonymous, and the design conforms to UNESCO's human-centred and data-protection principles (Miao & Holmes, 2023).

Data Analysis and Interpretation

Preliminary Analysis and Patterns of Use

Table 3 reports descriptive statistics and normality diagnostics. All skewness and kurtosis values fall within ±1, satisfying conventional criteria for univariate normality; Shapiro-Wilk tests reached significance for three variables, but at $N = 420$ the test is highly sensitive to trivial departures, so parametric procedures were retained.

Table 3
Descriptive Statistics and Normality Diagnostics (N = 420)

Variable	M	SD	Min	Max	Skew	Kurt	Shapiro W
AI Usage Intensity	3.84	0.68	1.57	5.00	-0.36	-0.06	.981
Cognitive Offloading	3.61	0.76	1.63	5.00	-0.29	-0.30	.985
Critical Thinking	3.28	0.74	1.19	5.00	-0.06	-0.44	.995
Metacognitive Self-Reg.	3.35	0.72	1.15	5.00	-0.08	-0.31	.995
AI Literacy	3.43	0.74	1.06	5.00	-0.09	-0.12	.994

Nrityanjali Multidisciplinary International Research Journal
Volume 2, Issue 2, April -June 2026
E-ISSN: 3108-1452
NMIRJ
Website: https://nrityanjaliresearchjournal.com/home/ Email: info@nrityanjaliresearchjournal.com

Perceived Benefit	3.89	0.65	2.14	5.00	-0.30	-0.45	.982
Academic Performance	7.42	0.95	4.60	9.91	0.02	-0.10	.997

Note. Constructs are scored on a five-point scale; academic performance is cumulative grade point average on a ten-point scale.

Figure 1 shows how often respondents turn to generative AI for academic work. Almost half the sample (48.8%) uses such tools at least several times a week, while 34.0% use them weekly, 14.3% monthly and only 2.9% never or rarely. AI use is no longer a minority behaviour at the margins of student practice; it is the modal condition of contemporary study.

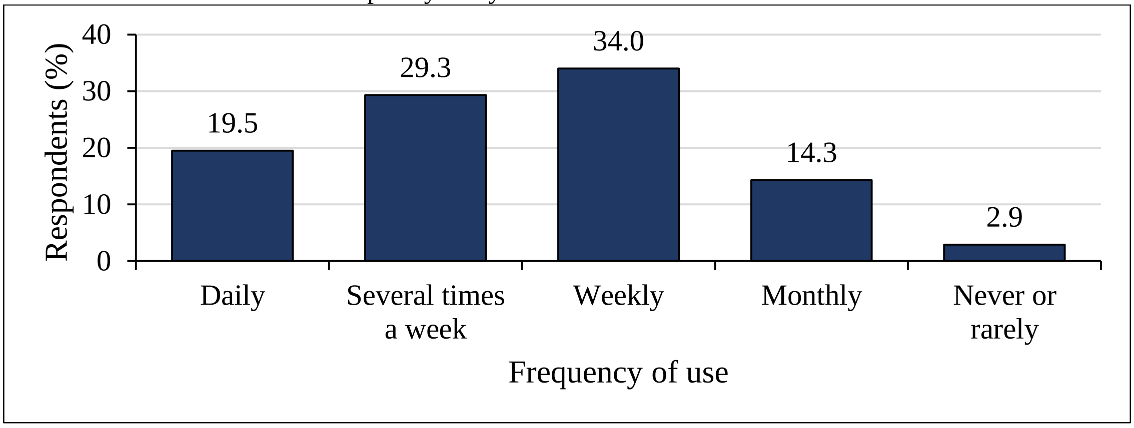

Figure 1: Distribution of Generative AI Usage Frequency among Respondents
Note. N = 420. Values are percentages of the total sample.

Purposes of use, recorded through a multiple-response item, cluster at the intellectual core of academic work rather than at its periphery: explaining difficult concepts (78.1%), drafting assignments and essays (71.4%), summarising reading material (64.5%), solving numerical problems (52.6%), language editing and translation (49.3%) and examination preparation (44.0%) all outrank mechanical tasks such as generating references (22.9%). Students are using AI to perform the thinking, not to tidy the output.

4.2 Relationship between Study Variables

Pearson correlations are reported in Table 4. AI usage is strongly positively associated with cognitive offloading ($r = .685$, $p < .001$) and strongly negatively with critical thinking ($r = -.734$, $p < .001$); offloading is itself negatively associated with critical thinking ($r = -.728$) and academic performance ($r = -.520$, both $p < .001$). H_{01} and H_{02} are rejected. Two findings complicate any simple condemnation of the technology: perceived benefit rises with usage ($r = .665$, $p < .001$) even as measured critical thinking falls, and AI literacy is unrelated to critical thinking ($r = .027$, $p = .587$) though positively related to performance ($r = .301$, $p < .001$) — knowing how AI works is not the same as being protected from it.

Table 4: Pearson Correlation Matrix among Study Variables (N = 420)

Variable	1	2	3	4	5	6	7
1. AI Usage	—						
2. Cognitive Offloading	.685**	—					
3. Critical Thinking	-.734**	-.728**	—				
4. Self-Regulation	-.532**	-.516**	.717**	—			
5. AI Literacy	.125*	.011	.027	.261**	—		
6. Perceived Benefit	.665**	.461**	-.446**	-.258**	.285**	—	
7. Academic Performance	-.467**	-.520**	.611**	.672**	.301**	-.253**	—

Note. *p < .05. **p < .01 (two-tailed). Variable numbers correspond to the row labels.

Figure 2 plots the individual-level relationship with a fitted linear trend. The scatter shows a consistent downward gradient across the full range of usage rather than a threshold beyond which harm begins, which argues against the intuitive policy of setting a permitted usage ceiling. Figure 3 displays the

Nrityanjali Multidisciplinary International Research Journal
Volume 2, Issue 2, April -June 2026
E-ISSN: 3108-1452
NMIRJ
Website: https://nrityanjaliresearchjournal.com/home/ Email: info@nrityanjaliresearchjournal.com

same relationship by decile: critical thinking falls from 4.19 in the lowest decile to 2.34 in the highest while cognitive offloading rises from 2.66 to 4.53, the two curves crossing between the fourth and fifth deciles — a point that may be read as the threshold at which the learner ceases to be the principal agent of the cognitive task.

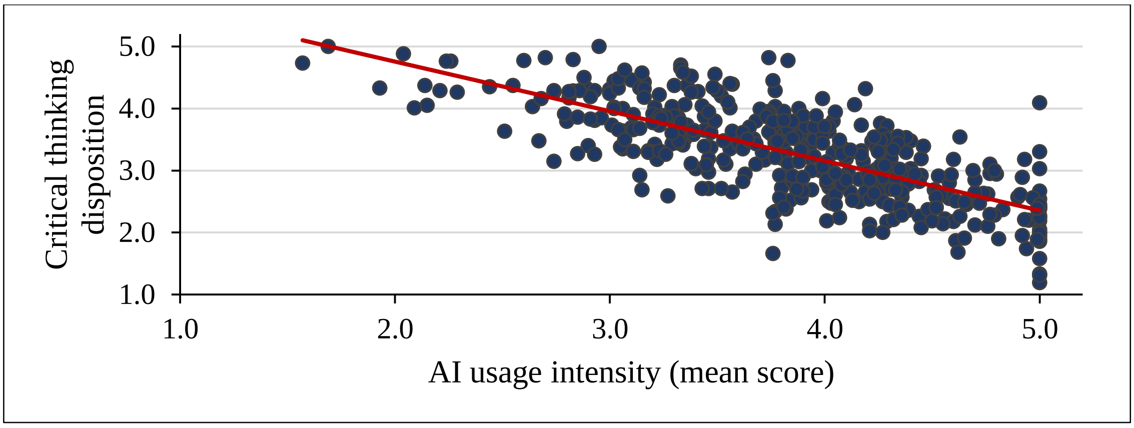

Figure 2: Scatterplot of AI Usage Intensity against Critical Thinking Disposition
Note. Each marker represents one respondent (N = 420); the fitted line is an ordinary least squares regression, r = -.734, p < .001.

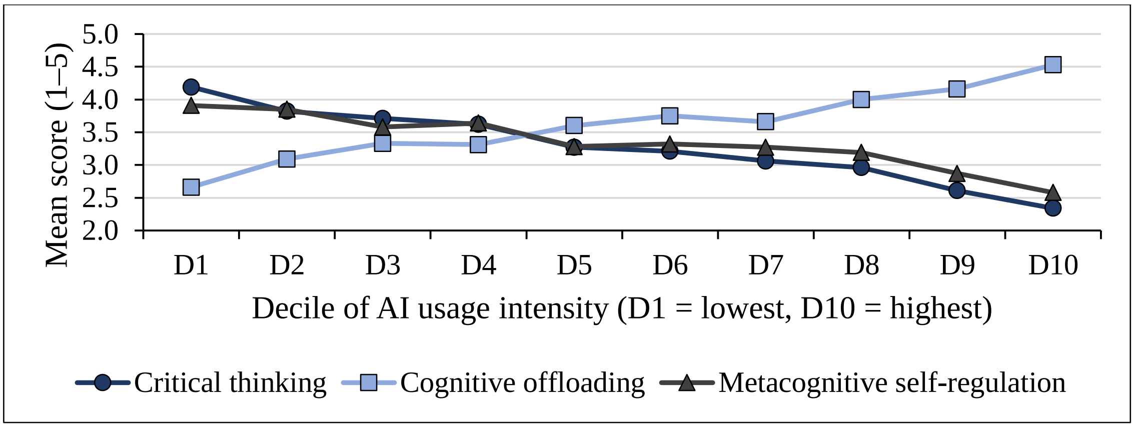

Figure 3: Dose–Response Pattern across Deciles of AI Usage Intensity
Note. Each decile contains 42 respondents.

Differences across Demographic and Usage Groups

Independent-samples t tests found no significant gender differences: AI usage, t(418) = 1.241, p = .215, d = 0.121; cognitive offloading, t(418) = 0.403, p = .687; critical thinking, t(418) = -1.248, p = .213, d = -0.122; academic performance, t(418) = -0.960, p = .338. Levene's test confirmed homogeneity of variance and all effect sizes were negligible, so H_{04} is retained — a useful null result indicating the phenomenon is generational and technological rather than gendered.

One-way ANOVAs were then run with AI usage group, formed by tertile split, as the independent variable. Table 5 reports significant differences on every dependent variable; the effect on critical thinking is very large, F(2, 417) = 172.01, p < .001, η^2 = .452, so usage group accounts for 45.2% of the variance. Tukey post hoc comparisons confirmed all three groups differ significantly on critical thinking, offloading, self-regulation and academic performance (all p < .01), and H_{03} is rejected. Supplementary analyses found no differences by level of study, F(2, 417) = 2.17, p = .115, or discipline, F(3, 416) = 0.26, p = .852. Figure 4 displays the group profile.

Table 5: One-Way ANOVA of Study Variables by AI Usage Group

Variable	**Low M (SD)**	**Moderate M (SD)**	**High M (SD)**	**F(2, 417)**	**p**	**η²**

Nrityanjali Multidisciplinary International Research Journal
Volume 2, Issue 2, April -June 2026
E-ISSN: 3108-1452
NMIRJ
Website: https://nrityanjaliresearchjournal.com/home/ Email: info@nrityanjaliresearchjournal.com

Critical Thinking	3.88 (0.54)	3.30 (0.56)	2.66 (0.55)	172.01	<.001	.452
Cognitive Offloading	3.06 (0.67)	3.56 (0.57)	4.21 (0.53)	134.51	<.001	.392
Metacognitive Self-Reg.	3.75 (0.61)	3.37 (0.67)	2.93 (0.62)	59.75	<.001	.223
Perceived Benefit	3.40 (0.57)	3.94 (0.49)	4.34 (0.50)	113.19	<.001	.352
Academic Performance	7.85 (0.89)	7.51 (0.86)	6.90 (0.85)	43.88	<.001	.174
AI Literacy	3.34 (0.73)	3.40 (0.77)	3.57 (0.72)	3.64	.027	.017

Note. n = 140 per group. η^2 values of .01, .06 and .14 denote small, medium and large effects respectively.

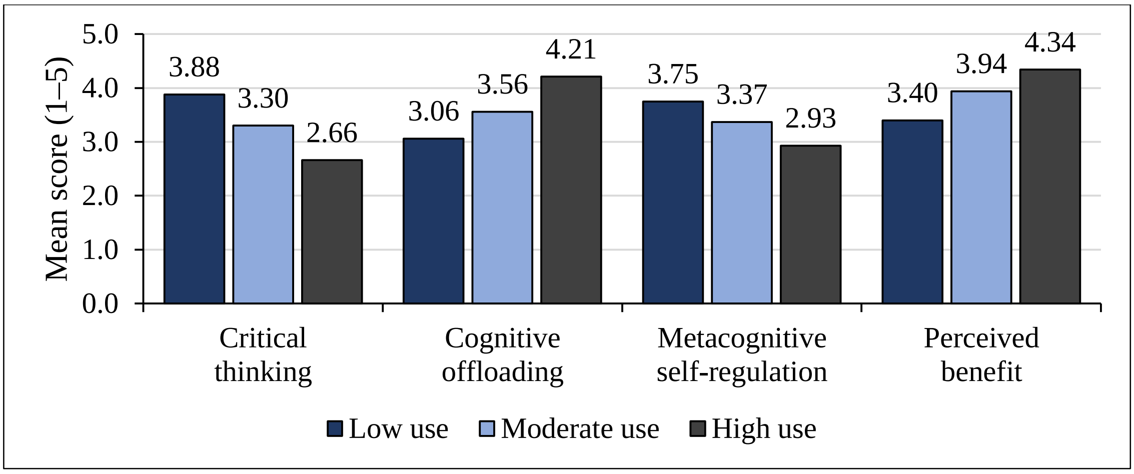

Figure 4: Mean Construct Scores across Low, Moderate and High AI Usage Groups
Note. n = 140 in each usage group; all between-group differences are significant at p < .001.

Predictors of Critical Thinking Disposition

A simultaneous multiple regression with critical thinking as criterion was significant, $F(4, 415) = 291.73$, $p < .001$, explaining 73.8% of variance (adjusted $R^2 = .735$, SE = 0.380); Durbin-Watson of 1.987 indicates independence of residuals and all variance inflation factors were below 2.3, so H_{05} is rejected.

Table 6
Multiple Regression Predicting Critical Thinking Disposition

Predictor	B	SE B	β	t	p	VIF
(Constant)	4.358	0.211	—	20.703	<.001	—
AI Usage Intensity	-0.324	0.041	-.297	-7.939	<.001	2.218
Cognitive Offloading	-0.306	0.035	-.313	-8.804	<.001	2.004
Metacognitive Self-Reg.	0.420	0.034	.408	12.318	<.001	1.733
AI Literacy	-0.039	0.027	-.040	-1.439	.151	1.198

Note. R = .859, $R^2 = .738$, adjusted $R^2 = .735$, $F(4, 415) = 291.73$, $p < .001$, Durbin–Watson = 1.987.

Three findings deserve emphasis. Metacognitive self-regulation is the strongest predictor ($\beta = .408$) and the only substantial positive one; AI usage and offloading make independent negative contributions of comparable magnitude, so volume and manner of use are distinct risks; and AI literacy is not significant ($\beta = -.040$, $p = .151$), meaning technical familiarity confers no protection against cognitive cost. A parallel regression on academic performance explained 53.3% of variance, $F(5, 414) = 94.51$, $p < .001$; once cognitive and regulatory variables were controlled, AI usage ceased to predict performance ($\beta = -.046$, $p = .391$) while self-regulation ($\beta = .381$), AI literacy ($\beta = .204$), critical thinking ($\beta = .180$) and offloading ($\beta = -.163$) remained significant. The bivariate association between heavy AI use and poorer grades is thus not a direct effect of the tool but operates through what the tool does to the learner's cognitive processes.

Mediation Analysis and Perceived Dependence

The causal steps procedure with a bias-corrected bootstrap of 5,000 resamples tested whether offloading transmits the effect of AI usage on critical thinking (Table 7). The total effect was significant (B = -0.800, $p < .001$), as were the paths from usage to offloading (B = 0.766) and from offloading to critical thinking (B = -0.414). Entering the mediator reduced the direct effect to -0.483 while it remained significant, establishing partial mediation. The indirect effect of -0.317 was significant by Sobel test (z = -9.174, $p < .001$) and the bootstrap interval [-0.385, -0.248] excluded zero, so offloading accounts for 39.6% of the total effect and H_{06} is rejected.

Table 7: Mediation of the AI Usage – Critical Thinking Relationship by Cognitive Offloading

Path	Description	B	SE	t / z	p
a	AI usage → Cognitive offloading	0.766	0.040	19.202	<.001
b	Offloading → Critical thinking	-0.414	0.040	-10.444	<.001
c	Total effect: usage → thinking	-0.800	0.036	-22.080	<.001
c′	Direct effect controlling mediator	-0.483	0.044	-10.906	<.001
a × b	Indirect (mediated) effect	-0.317	0.035	-9.174	<.001

Note. Bootstrap (5,000 resamples) 95% CI for the indirect effect = [-0.385, -0.248]; proportion of total effect mediated = 39.6%; model $R^2 = .634$.

A chi-square test examined the association between usage group and self-reported inability to complete academic work without AI. It was highly significant, $\chi^2(2, N = 420) = 134.71$, $p < .001$, Cramér's V = .566, with dependence rising from 17.1% of light users to 52.9% of moderate and 86.4% of heavy users; H_{07} is rejected. No association was found between discipline and dependence, $\chi^2(3, N = 420) = 2.96$, p = .398. Asked to judge the net effect of AI on their own thinking, 40.0% reported strengthening and 38.4% weakening — a near-even split that, read alongside Table 4, illustrates the difficulty of self-assessment here: those most likely to report benefit are, on the measured evidence, those experiencing the greatest cognitive substitution.

Discussion

The Central Paradox

The results reproduce, within one sample, the contradiction running through the wider literature. Perceived benefit rises steeply with usage, reaching 4.34 among heavy users, yet critical thinking falls from 3.88 to 2.66 across the same gradient. The trends are not contradictory once it is recognised that perceived benefit tracks task completion while critical thinking tracks the process by which the task is completed. This reconciles the opposed literatures: Deng et al. (2025), Liu et al. (2025) and Ma and Zhong (2025) measure performance on tasks completed with AI available, whereas Gerlich (2025), Kosmyna et al. (2025) and Fan et al. (2025) measure what the learner retains afterwards.

Cognitive Offloading as the Operative Mechanism

The mediation analysis identifies where the damage occurs. Offloading carries 39.6% of the total effect, with a bootstrap interval comfortably excluding zero. The estimate is consistent with Gerlich (2025), whose correlations were of comparable magnitude (r = .72 and r = -.75, against .685 and -.728 here), and with the dual-pathway model of Ma et al. (2026): three independent samples from different national contexts converging within .05 of one another is a meaningful replication. Because mediation is partial, interventions targeting offloading alone will not eliminate the association — but the harmful component of AI use is substitution rather than exposure, a considerably more tractable target.

Self-Regulation as the Protective Factor

The most actionable result is the contrast between metacognitive self-regulation and AI literacy. Self-regulation is the strongest predictor of both critical thinking (β = .408) and academic performance (β = .381), whereas AI literacy fails to predict critical thinking at all (β = -.040, p = .151). Understanding how a language model works evidently does not translate into the discipline of deciding when not to use one. This aligns with Shi et al. (2025) and qualifies current competency frameworks: the present data suggest a fifth

Nrityanjali Multidisciplinary International Research Journal
Volume 2, Issue 2, April -June 2026
E-ISSN: 3108-1452
NMIRJ Website: https://nrityanjaliresearchjournal.com/home/ Email: info@nrityanjaliresearchjournal.com

dimension on regulatory self-management during AI-assisted work should join the four specified by UNESCO (Miao et al., 2024).

The Perception Gap and Theoretical Implications

Heavy users report the highest perceived benefit while recording the lowest measured critical thinking, self-regulation and grades; Lee et al. (2025) documented the same asymmetry among knowledge workers. Fluency is being mistaken for understanding, and because students cannot detect the deficit through introspection, self-report can trigger neither intervention nor evaluation. For cognitive load theory, the load removed by generative AI is germane rather than extraneous: reduced mental effort is no clean gain if the effort removed was the effort that produced learning. For distributed cognition, the assembly outperforms the unaided human only while the human retains the evaluative function, and Figure 3 locates that transition around the fourth to fifth decile. For self-regulated learning theory, regulation becomes the moderator determining whether a technology functions as scaffold or substitute.

Practical Implications

For institutions, the evidence argues against both permissive and prohibitionist policies and in favour of process-visible assessment: since the harm lies in substitution rather than exposure, students should submit reasoning traces, prompt histories and verification records alongside final products. For teachers, Niloy et al. (2024) supply the caution that rote-oriented pedagogy was the strongest driver of chatbot substitution, meaning tasks reproducible by a language model will be reproduced by one. For students, the lesson is deliberate sequencing — attempting generation before consultation. For policymakers, Singh et al. (2026) identify the absence of institutional guidance as the most salient barrier reported by Indian students and teachers alike.

Conclusion, Limitations and Recommendations

Conclusion

The question posed in the title admits no categorical answer, and the data indicate that asking it categorically is itself the error. Artificial intelligence in education is neither a threat to cognitive ability nor a better learning tool as a fixed property of the technology. Across 420 students the same tools were associated with the highest perceived benefit and the lowest measured critical thinking, and what separated the outcomes was the learner's regulatory relationship to the tool. Where students retained the evaluative function AI operated as an amplifier; where they surrendered it AI operated as a substitute. Three conclusions follow: cognitive offloading is a measurable mechanism rather than a metaphor; metacognitive self-regulation rather than AI literacy is the protective factor institutional effort should target; and students cannot see the effect in themselves, so curriculum and assessment design must carry the burden of the response.

Limitations of the Study

Five limitations qualify these conclusions. The cross-sectional design establishes association rather than causation, and the possibility that students with weaker critical thinking are drawn to heavier AI use cannot be excluded. Critical thinking was measured as a self-reported disposition rather than by a performance instrument such as the Watson-Glaser test, and the perception gap in section 5.4 suggests self-report may understate the effect. Academic performance was self-reported and subject to recall and social desirability bias. The sample does not support national generalisation. Finally, common method variance is a plausible concern, since all constructs were measured through one self-administered instrument at a single time point.

Recommendations

Institutions should replace blanket permission or prohibition with disclosure-based policies requiring students to document how AI was used, and should redesign assessment so that the reasoning process rather than the finished artefact carries the marks. Teachers should rewrite assessment tasks a language model can complete unaided, embed deliberate unaided phases in coursework, and reward the identification of machine error. Students should adopt a generate-then-consult sequence and periodically test retention without the tool. Policymakers should extend national frameworks beyond technical AI literacy to mandate metacognitive and regulatory competencies (Miao et al., 2024).

Scope for Further Research

Nrityanjali Multidisciplinary International Research Journal
Volume 2, Issue 2, April -June 2026
E-ISSN: 3108-1452
NMIRJ Website: https://nrityanjaliresearchjournal.com/home/ Email: info@nrityanjaliresearchjournal.com

Longitudinal designs tracking one cohort across several semesters would resolve the direction of causation, and experimental work manipulating the sequencing of AI access within a task would test the generate-then-consult recommendation directly. Objective critical thinking measures combined with process data of the kind used by Kosmyna et al. (2025) would establish whether self-report understates the effect, and intervention studies comparing AI literacy training against metacognitive regulation training would supply the causal test of the comparative claim advanced here.

References

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544

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

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task (arXiv:2506.08872) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.08872

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). 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 (Article 1121, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P.-M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. npj Science of Learning, 10, Article 29. https://doi.org/10.1038/s41539-025-00320-7

Liu, X., Guo, B., He, W., & Hu, X. (2025). Effects of generative artificial intelligence on K-12 and higher education students' learning outcomes: A meta-analysis. Journal of Educational Computing Research, 63(5), 1249–1291. https://doi.org/10.1177/07356331251329185

Ma, N., & Zhong, Z. (2025). A meta-analysis of the impact of generative artificial intelligence on learning outcomes. Journal of Computer Assisted Learning, 41(5), Article e70117. https://doi.org/10.1111/jcal.70117

Ma, Y., Zhang, Z., & Liu, C. (2026). The double-edged sword effect of GenAI assistance on university students' academic performance: Evidence from China. Education and Information Technologies, 31(3), 675–700. https://doi.org/10.1007/s10639-025-13850-9

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., … Oak, S. (2025). Artificial Intelligence Index Report 2025. Stanford Institute for Human-Centered AI. https://doi.org/10.48550/arXiv.2504.07139

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Miao, F., Shiohira, K., & Lao, N. (2024). AI competency framework for students. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391105

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

Nie, X., Tian, Y., Liu, M., Wu, D., & Guo, Y. (2025). The impact of generative artificial intelligence on students' higher order thinking: Evidence from a three-level meta-analysis. Education and Information Technologies, 30(17), 25359–25390. https://doi.org/10.1007/s10639-025-13735-x

Niloy, A. C., Hafiz, R., Hossain, B. M. T., Gulmeher, F., Sultana, N., Islam, K. F., Bushra, F., Islam, S., Hoque, S. I., Rahman, M. A., & Kabir, S. (2024). AI chatbots: A disguised enemy for academic integrity? International Journal of Educational Research Open, 7, Article 100396. https://doi.org/10.1016/j.ijedro.2024.100396

Shi, J., Liu, W., & Hu, K. (2025). Exploring how AI literacy and self-regulated learning relate to student writing performance and well-being in generative AI-supported higher education. Behavioral Sciences, 15(5), Article 705. https://doi.org/10.3390/bs15050705

Singh, A., Bajpai, A., & Singh, S. (2026). Perceptions, challenges and ethical concerns of ChatGPT use in Indian higher education institutions. Discover Education, 5(1), Article 283. https://doi.org/10.1007/s44217-026-01300-8

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7