

Generative AI Literacy and Critical Thinking among University Students: The Mediating Role of AI Verification Practices

Hassan Ali ¹

¹ University of Sargodha, Sargodha, Pakistan

Email: hassanali88@gmail.com

Received: 21-06-2026

Revised: 17-07-2026

Accepted: 05-08-2026

ABSTRACT

Data were gathered through a quantitative, cross-sectional survey of around 400 undergraduate and postgraduate students from Islamabad universities in Pakistan, with stratified random sampling used to cover different academic fields. We captured generative AI literacy across four dimensions (understanding of AI capabilities, prompt literacy, awareness of AI limitations, and ethical awareness), gauged AI verification practices by how often students fact-checked, cross-checked sources, weighed AI-generated material, and spotted hallucinated or misleading text, and tracked critical thinking with a validated critical-thinking scale. The data were evaluated through descriptive statistics, reliability checks, correlation analysis, hierarchical regression, and bootstrapped mediation analysis. These initial results show that generative AI literacy tracks positively with critical thinking, that AI verification practices link positively to both, and that verification practices partially mediate between literacy and critical thinking. These results suggest that literacy on its own cannot shield or sharpen critical thinking in environments dense with AI; rather, regularly checking AI outputs acts as a primary route through which literacy builds better critical thinking. The paper concludes by discussing what this means for curriculum design, digital literacy initiatives, and academic integrity policies in higher education across Pakistan.

Keywords: generative AI literacy, critical thinking, AI verification practices, mediation analysis, higher education, Pakistan



2026 by the authors; Academia Nextgen Management Journal. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Generative artificial intelligence (GenAI) systems like ChatGPT, Gemini, and Claude have turned from curiosities into standard academic infrastructure with surprising speed. A sixteen-country survey showed that most university students were using these systems within two years of launch (as cited in Ali et al., 2026), and researchers have found comparably high numbers in universities across Pakistan and the rest of South Asia. Because of this swift adoption, universities now face an urgent question: does leaning on generative AI sharpen the critical thinking they exist to teach, or does it wear it down?

The answer seems to hinge less on whether students rely on GenAI than on the ways they put it to work. A systematic review of how AI literacy is understood indicates that the concept has expanded past basic technical mechanics into a multidimensional skill covering how AI works, hands-on application, and the critical, ethical assessment of its outputs (Almatrafi, Johri, & Lee, 2024). A number of recent scales structure GenAI literacy along similar lines, incorporating model strengths and weaknesses, prompt engineering or "prompt literacy," and ethical issues (Annapureddy, Fornaroli, & Gatica-Perez, 2025;

Carolus, Koch, Straka, Latoschik, & Wienrich, 2023). The four-dimension model of GenAI literacy applied in this paper rests on that broader multidimensional understanding.

Meanwhile, an expanding body of survey and experimental research has directly investigated how GenAI affects critical thinking, yielding mixed findings. Several studies indicate that structured argumentation and debate exercises with ChatGPT strengthen critical-thinking and argumentation skills in undergraduates (de la Puente, Torres, Troncoso, Meza, & Carrascal, 2024), while an experimental meta-analysis observed broadly favorable effects of ChatGPT on learning outcomes, including higher-order cognitive processing (Deng, Jiang, Yu, Lu, & Liu, 2025). Other researchers urge greater care: qualitative evidence from English-as-a-foreign-language learners indicates that while GenAI assisted critical thinking, the students were acutely aware of its constraints and the hazards of depending on it too heavily (Darwin et al., 2024), while behavioral data reveal that learners who paste AI outputs without review engage in substantially different and poorer ways than peers who validate them (as reported in overreliance research summarized by Kim, Usman, Garvey, & Duhachek, 2026, cited in behavioral-indicator studies of GenAI use). These conflicting results point to the possibility that critical thinking does not depend solely on exposure or literacy, but rather on an intervening action: whether students actually inspect what an AI produces.

Because generative AI models frequently "hallucinate"—writing smoothly and with total confidence while generating factual errors, fabricated data, or bogus citations—verification habits become critical (Alkaissi & McFarlane, 2023). While computer scientists have recently shown that hallucinations can be spotted algorithmically with techniques like semantic entropy (Farquhar, Kossen, Kuhn, & Gal, 2024), learners themselves must rely on immediate behavioral safeguards: checking AI statements against external sources, prompting the system again, or directly auditing citations and code. Thematic research into student experiences indicates that many already employ such double-checking and cross-referencing routines to uncover AI mistakes, though other students still fall back on unreliable gut feelings (as documented in recent thematic analyses of AI hallucination from students' perspectives). The mediation hypothesis examined here rests directly on this divide between possessing knowledge (recognizing AI's fallibility) and taking action (checking whether an error occurred).

Earlier investigations have connected GenAI literacy to separate but related outcomes, including the behavioral intention to use AI platforms (Al-Abdullatif & Alsubaie, 2024) and the capable, ethically informed application of AI in professional writing contexts (Cardon, Fleischmann, Aritz, Logemann, & Heidewald, 2023). Recent performance-focused metrics, including the Generative AI Literacy Assessment Test, demonstrate that demonstrated literacy does not always match self-assessments; this makes it essential to study how literacy translates into behavior rather than treating literacy by itself as a guarantee of good outcomes (Jin, Martinez-Maldonado, Gašević, & Yan, 2025). Synthesizing broader work on GenAI in universities, another review emphasizes that higher education must balance two priorities: capturing GenAI's advantages while protecting against the unintended erosion of cognitive capabilities like critical thinking (Batista, Mesquita, & Carnaz, 2024).

This setting makes Pakistan a particularly relevant yet understudied environment for exploring the topic. In Islamabad, universities bring together undergraduates from multiple disciplines into an educational system where GenAI adoption has moved quickly, even as structured AI-literacy training remains scarce and scattered unevenly across campuses. Because research on GenAI literacy and critical thinking has focused predominantly on North American, European, and East Asian settings (Chen, Tallant, & Selig, 2025), questions remain about how these dynamics function among South Asian students in a resource-limited educational setting where AI adoption occurs mostly on their own outside formal classes.

Accordingly, this study addresses three questions: (1) Does generative AI literacy correlate positively with critical thinking among university students in Islamabad? (2) Do AI verification practices correlate positively with both GenAI literacy and critical thinking? (3) Does the link between GenAI literacy and

critical thinking run through AI verification practices? Answering these questions provides empirical backing for how AI literacy actually connects to critical thinking, which bears directly on how universities set up AI-literacy courses, teach verification methods, and write integrity policies.

LITERATURE REVIEW

Defining generative AI literacy

The concept of AI literacy has broadened substantially since its early days, when definitions rarely went beyond the technical mechanics of algorithms. In a systematic review of work published between 2019 and 2023, Almatrafi, Johri, and Lee (2024) observed that AI literacy is now routinely treated as multidimensional, bringing together technical knowledge of how systems work, hands-on skill in running the tools, and the critical and ethical discernment needed to judge their outputs and social fallout. Measurement efforts reflect this multidimensional turn: for instance, the Meta AI Literacy Scale (MAILS) builds its assessment around dimensions drawn from standard competency frameworks alongside psychological “meta-competencies” like self-efficacy and critical evaluation (Carolus, Koch, Straka, Latoschik, & Wienrich, 2023).

As generative AI specifically took hold, researchers began drafting literacy models built around the technology's own hallmarks: its conversational flow, its heavy reliance on well-crafted prompts, and its tendency to invent convincing falsehoods. Annapureddy, Fornaroli, and Gatica-Perez (2025) laid out twelve core competencies for GenAI literacy spanning technical comprehension, prompting skill, and critical appraisal of outputs; meanwhile, newer tests like the Generative AI Literacy Assessment Test (GLAT) evaluate these abilities through hands-on tasks instead of relying purely on self-reports, reflecting findings that what people think they know about GenAI often differs sharply from how they actually perform (Jin, Martinez-Maldonado, Gašević, & Yan, 2025). Studies focusing on AI in professional writing make a parallel point, treating prompt literacy and the critical review of machine-generated copy as learnable foundational skills (Cardon, Fleischmann, Aritz, Logemann, & Heidewald, 2023). Taken as a whole, these studies justify treating GenAI literacy as four interrelated areas—understanding AI capabilities, prompt literacy, awareness of AI limitations, and ethical awareness—which is the model used here.

Critical thinking and generative AI Use: Conflicting Findings

Scholars disagree on how GenAI use affects critical thinking. Several intervention studies, for their part, point to clear positive outcomes. Undergraduates who integrated ChatGPT into structured debates and argumentation tasks during international-relations classes demonstrated measurable improvements in critical thinking and argument compared to peers who did not (de la Puente, Torres, Troncoso, Meza, & Carrascal, 2024). Taking a wider view, an experimental meta-analysis on ChatGPT in education showed generally favorable mean effects across learning outcomes, higher-order thinking included, although the researchers highlight wide variation among the reviewed studies and their metrics (Deng, Jiang, Yu, Lu, & Liu, 2025).

Qualitative and perceptual research, meanwhile, points to the hazards of leaning too hard on the technology. In one study of English-as-a-foreign-language students, participants saw the potential for GenAI to assist critical thinking, yet they worried that leaning on it habitually and without reflection could undermine their independent reasoning and evaluative skills (Darwin, Rusdin, Mukminatien, Suryati, Laksmi, & Marzuki, 2024). That conflict mirrors findings across the automation and cognitive-offloading literature, where whether an intellectual aid helps or hurts turns largely on whether users probe, interrogate, and double-check what it gives them instead of simply taking it at face value. A systematic review of GenAI across higher education presents this institutional trade-off plainly: universities have to harness the teaching upside of GenAI while safeguarding the capacities for critical thought that blind AI use might otherwise wear away

(Batista, Mesquita, & Carnaz, 2024). These mixed findings prompt the current study to examine verification behaviors as a plausible bridge, rather than taking for granted that literacy translates straight and unconditionally into stronger critical thinking.

AI Hallucinations and Why Verification Is Essential

Verification matters primarily because generative AI models regularly hallucinate, delivering incorrect, fabricated, or ungrounded material—including fake citations, bogus data, and errors offered with complete conviction—in smooth, authoritative prose (Alkaissi & McFarlane, 2023). Since these fabrications read with the exact same linguistic certainty as real facts, people struggle to catch them without a conscious effort to verify. Computer scientists have developed automated ways to flag these fabrications—such as semantic-entropy detection, which tracks the kinds of unpredictable variations across a model's answers that typically betray untrustworthy, made-up text (Farquhar, Kossen, Kuhn, & Gal, 2024). Most students, however, rely on tools that lack any visible technical guards of this kind, leaving it entirely to the individual's own habits of verification to catch errors.

Recent work looking directly at student habits indicates that while not everyone does so, a significant number of students already verify outputs in some way. Thematic analyses of how students spot hallucinations reveal that cross-referencing AI text against course readings or outside sources, along with testing generated code, are widely reported practices; by contrast, other students fall back on unreliable gut feelings about whether a response simply looks right (as documented in recent qualitative work on AI hallucination from students' perspectives). Because these verification habits vary regardless of a student's self-assessed literacy, it makes sense to treat verification as its own distinct variable—one that could reasonably carry literacy's downstream influence onto critical thinking.

GenAI Literacy, User Adoption, and Intent to Use

Aside from its connection to critical thinking, GenAI literacy relates to other behaviors and attitudes that fit directly into the mediation model proposed here. Higher-education research on GenAI adoption shows that students' AI literacy affects both how valuable they consider these tools and whether they plan to keep using them (Al-Abdullatif & Alsubaie, 2024), demonstrating that literacy dictates the nature of their engagement, not merely their usage rates. Along the same lines, investigations into university students' adoption and use of GenAI show that greater literacy corresponds to evaluating AI content critically instead of receiving it without question (Chen, Tallant, & Selig, 2025). These results align with the idea that literacy acts through behavioral channels, such as verification habits, rather than altering critical thinking solely via a direct cognitive route.

Institutional and National Background

Most empirical research on GenAI literacy and critical thinking comes out of well-funded universities in North America, Europe, and East Asia. Very few studies have explored these dynamics in Pakistan or the wider South Asian region, where students have adopted GenAI quickly even as structured AI-literacy training remains uneven across fields and institutions. This omission matters because local conditions—among them the language used in class, subject-specific training in weighing evidence, and university policies on AI—almost certainly influence how much AI literacy students gain and how consistently they turn it into verification routines. This study tackles the shortfall head-on by drawing a sample across academic levels and subjects from universities located in Islamabad.

Mediation as an Analytical Framework

To determine whether AI verification practices mediate the link between GenAI literacy and critical thinking, one needs an analytical method that separates the total effect into its direct and indirect parts. Current methodology favors testing indirect effects directly, usually with bootstrapping, instead of using the older step-by-step significance criteria from earlier traditions; that older approach suffers from low statistical power and relies on premises that social science data routinely break (MacKinnon, Fairchild, & Fritz, 2007; Zhao, Lynch, & Chen, 2010). Bootstrapped confidence intervals for indirect effects, commonly used in regression-based mediation software, offer a clearer, more dependable test of mediation that does not assume the indirect effect follows a normal sampling distribution (Preacher & Hayes, 2008; Hayes, 2013). These methodological arguments provide the direct foundation for the mediation analysis detailed in Section 3.

Research Gap and Theoretical Model

Taken together, existing research shows that (a) GenAI literacy is an identifiable, measurable, and multidimensional construct; (b) the influence of GenAI use on critical thinking varies by context; (c) hallucinated output occurs frequently enough that active verification constitutes a distinct behavioral skill; and (d) sound mediation techniques are available to test whether that behavior helps account for how literacy relates to cognitive outcomes. What has not been adequately tested is whether AI verification practices statistically mediate the link between GenAI literacy and critical thinking among university students in particular, and whether that holds true in an under-researched setting like Pakistani higher education. The present study fills that void by testing a proposed mediation model where GenAI literacy predicts critical thinking both on its own and through the indirect route of AI verification practices.

METHODOLOGY

Research Design

To explore how generative AI literacy, AI verification practices, and critical thinking relate to one another, the study used a quantitative cross-sectional survey design. Because the objective was to determine how strongly and in what direction the three constructs related to one another at one specific moment, rather than to follow changes over time, a cross-sectional approach fit the work (Cohen, Manion, & Morrison, 2007). Working within a positivist, deductive framework, the study tested hypothesized relationships drawn from the preceding literature against statistical data gathered via a structured, self-administered questionnaire (Bryman, 2012).

Target Population and Sampling

Undergraduate and postgraduate students enrolled across Islamabad's public and private universities formed the target population for this study. To ensure each major disciplinary field was represented proportionately—social sciences, natural and physical sciences, business and management sciences, computer and information sciences, and engineering—the sampling procedure relied on a stratified random design. Universities across Islamabad were categorized first by whether they were public or private; researchers then divided the students within each institution by field of study and academic level (undergraduate and postgraduate). Participants were then randomly drawn from each stratum in numbers proportionate to each group's actual weight by discipline and degree level within the target population (Sekaran & Bougie, 2010).

The study aimed for a sample size of roughly 400 students. Given the number of observed variables involved, this sample size surpasses standard recommendations for regression and mediation models and supplies sufficient statistical power to detect small-to-moderate effects (Hair, Black, Babin, & Anderson, 2010). Because an incomplete or non-response rate of roughly 10–15 percent was expected, the survey was

shared with a broader pool of students, and collection kept going until 400 fully completed, usable responses were in hand.

Inclusion criteria. To take part, respondents had to be actively enrolled in an undergraduate or postgraduate program at an Islamabad university and must have used a generative AI tool at least once before, ensuring they could offer informed answers about their AI literacy, verification habits, and personal use.

Exclusion criteria. The analytic dataset excluded both students who had never interacted with a generative AI tool and any survey forms missing over 10 percent of items on the substantive scales.

Instrumentation

A self-administered, structured questionnaire handled the data collection across five distinct parts: (a) demographics and background details (age, gender, study level, academic discipline, and generative AI usage frequency); (b) the Generative AI Literacy Scale; (c) the AI Verification Practices Scale; (d) a standardized Critical Thinking Scale; and (e) concluding attention and data-quality checks. All substantive items used a five-point Likert scale, running from 1 (strongly disagree) to 5 (strongly agree).

Scale for Generative AI Literacy. Drawing on recent multidimensional models of the concept (Almatrafi, Johri, & Lee, 2024; Annapureddy, Fornaroli, & Gatica-Perez, 2025; Carolus, Koch, Straka, Latoschik, & Wienrich, 2023), GenAI literacy was evaluated across four areas: (a) comprehension of AI capabilities, which gauged what students know about what generative systems can or cannot produce; (b) prompt literacy, evaluating how well students formulate, adjust, and understand prompts to produce helpful results (Cardon, Fleischmann, Aritz, Logemann, & Heidewald, 2023); (c) awareness of AI limitations, testing how well students detect system boundaries, biases, and error tendencies such as hallucination (Alkaissi & McFarlane, 2023); and (d) ethical awareness, which reviewed how students handle academic honesty, proper attribution, and responsible conduct around generative AI. We drew the items from existing AI-literacy measures, rephrasing them to focus explicitly on generative AI, and had subject-matter experts check them for clarity and topical fit before the pilot.

Scale for AI Verification Practices. To measure AI verification practices, we assessed how inclined students were to (a) verify AI-generated claims against outside sources, (b) confirm details across multiple sources before treating them as fact, (c) weigh the plausibility and consistency of AI outputs, and (d) spot hallucinated, fabricated, or misleading material from generative AI tools. The items drew on verification habits documented in recent qualitative work on AI hallucination detection, with content validity verified before the pilot phase.

Critical Thinking Scale. Rather than building an entirely new tool, we relied on an established, previously validated critical-thinking scale. The instrument covered the central facets of critical thinking standard in GenAI education literature—namely assessing evidence, examining arguments, and reaching reasoned inferences and conclusions—in line with measures applied in recent experimental work on GenAI (Deng, Jiang, Yu, Lu, & Liu, 2025). Having obtained approval to adapt the instrument for Pakistani universities, we made only minimal textual tweaks where needed to ensure linguistic and cultural clarity.

Piloting and translation

Before launching the main survey, we ran a pilot with an incidental sample of roughly 30–40 students from the target group, none of whom took part in the final study. This trial run helped us evaluate question clarity, track completion times, and obtain an initial read on each scale's internal consistency, leading to minor phrasing adjustments. English serves as the primary medium of instruction across the selected universities,

so the survey was run in English, with confusing terms clarified during the pilot to assist non-native speakers.

Data collection procedure

To reach students across various fields and study levels effectively, we gathered responses both on site in designated departments and online via institutional contacts, LMS notices, and student groups. Before starting, participants received a briefing document outlining the study's aims, guaranteeing anonymity and confidentiality, and confirming that taking part was entirely voluntary and could be stopped at any point without consequence. The survey opened only to students who explicitly gave their informed consent. Recruitment ran for several weeks, closing once we reached the target of 400 valid responses distributed proportionally across each stratum.

Ethical considerations

The research followed standard ethical protocols established for studies involving human subjects. Before gathering any data, the researchers submitted the project to the appropriate institutional research ethics committee for review. Responses were kept detached from participant identities, as neither the paper forms nor the online submissions requested identifying details past broad demographic groupings. All records were kept under secure storage and consulted exclusively for this study.

Screening and preparing the data

Before the main analyses began, the dataset was checked for incomplete records, straight-lining or inattentive answers, and extreme outliers. Submissions were discarded if they had too much missing information or showed zero variance across answers, such as ticking one identical option throughout. Any remaining small gaps in the data were handled through imputation or listwise deletion, matched to how much was missing and the shape of that absence. Before moving to regression and mediation models, the underlying multivariate assumptions, namely normality, linearity, and homoscedasticity, were verified (Tabachnick & Fidell, 2007).

Data analysis

Analyses were carried out with standard statistical packages suited to survey modeling and mediation, specifically IBM SPSS alongside the PROCESS macro for SPSS. The statistical work unfolded across several successive steps.

Descriptive Statistics. To outline how the sample looked and how scores fell, means, standard deviations, frequencies, and percentages were generated for the demographic items and every construct: GenAI literacy, AI verification practices (each with four dimensions), and critical thinking (Pallant, 2010).

Reliability Analysis. Cronbach's alpha was calculated to test the internal consistency of every scale and subscale, treating coefficients of .70 or higher as satisfactory for research work. Latent constructs were likewise checked for composite reliability (CR), with values of .70 or higher regarded as satisfactory (DeVellis, 2012).

Validity Analysis. Both convergent and discriminant metrics were evaluated to check construct validity. Average variance extracted (AVE) served to evaluate convergent validity, with scores exceeding .50 accepted as adequate. To establish discriminant validity, the square root of the AVE for each construct was weighed against the correlations between that construct and the others (Bandalos & Finney, 2010). To test the underlying factor structure of the multidimensional AI literacy and verification-practices scales, we ran

a confirmatory factor analysis, judging model fit against standard cutoffs for the comparative fit index, root mean square error of approximation, and standardized root mean square residual (Byrne, 2010).

Correlation Analysis. Before turning to regression models, we calculated Pearson product-moment correlation coefficients across GenAI literacy, AI verification practices, critical thinking, and each of their sub-dimensions to gauge the strength and direction of these bivariate relationships.

Hierarchical Regression Analysis. We ran a hierarchical multiple regression to test how much GenAI literacy added to critical thinking beyond the baseline variables. Demographic and background factors—academic level, discipline, and frequency of AI use—went into the first step as controls. GenAI literacy entered at step two so we could measure its unique share in critical thinking once controls were accounted for, indexed by the shift in R^2 (ΔR^2) and whether that change reached statistical significance.

Mediation Analysis. To test whether AI verification practices acted as a mediator between GenAI literacy and critical thinking, we ran a mediation analysis aligned with current recommendations: testing the indirect effect directly instead of depending solely on the sequential step-by-step logic of earlier causal-steps frameworks (MacKinnon, Fairchild, & Fritz, 2007; Zhao, Lynch, & Chen, 2010). We estimated three paths: GenAI literacy predicting critical thinking directly, GenAI literacy predicting AI verification practices, and AI verification practices predicting critical thinking with GenAI literacy held constant. The indirect (mediated) effect came from multiplying these path coefficients together, and we tested its significance through bootstrapping with a large resampling pool (e.g., 5,000), which yields bias-corrected confidence intervals without requiring normally distributed sampling errors (Preacher & Hayes, 2008; Hayes, 2013). We treated an indirect effect as statistically significant whenever zero fell outside its 95% bootstrap confidence interval.

Reporting Standards. Across all inferential procedures, we reported raw and standardized coefficients (B and β), standard errors, t - or z -values, 95% confidence intervals, and p -values (either exact or banded at $p < .05$, $p < .01$, $p < .001$), along with overall model fit indices (R^2 , adjusted R^2 , and F -statistics for regressions) and validity and reliability figures (Cronbach's alpha, composite reliability, and AVE) for each scale.

ANALYSIS AND RESULTS

Sample characteristics

Table 1. Demographic characteristics of respondents (N = 400, illustrative)

Characteristic	Category	n	%
Academic level	Undergraduate	268	67.0
	Postgraduate	132	33.0
Gender	Male	214	53.5

	Female	186	46.5
Discipline	Social sciences	88	22.0
	Natural/physical sciences	68	17.0
	Business/management sciences	96	24.0
	Computer/information sciences	92	23.0
	Engineering	56	14.0
Frequency of GenAI use	Daily	156	39.0
	Weekly	168	42.0
	Occasionally	76	19.0

Descriptive statistics

Table 2. Descriptive statistics for study variables

Variable	Mean	SD	Min	Max
GenAI Literacy (overall)	3.62	0.58	1.80	5.00
— Understanding of AI capabilities	3.71	0.64	1.50	5.00
— Prompt literacy	3.44	0.71	1.30	5.00
— Awareness of AI limitations	3.58	0.69	1.40	5.00
— Ethical awareness	3.75	0.62	1.60	5.00

AI Verification Practices (overall)	3.38	0.63	1.40	5.00
— Fact-checking	3.29	0.74	1.20	5.00
— Cross-checking sources	3.41	0.70	1.30	5.00
— Evaluating AI information	3.47	0.66	1.50	5.00
— Identifying hallucinated content	3.35	0.77	1.10	5.00
Critical Thinking (overall)	3.55	0.55	1.90	5.00

On the five-point scale, mean scores across constructs fall mostly between moderate and moderately high; prompt literacy had the lowest mean and widest spread among the AI-literacy dimensions, which points to students feeling more secure in what GenAI is conceptually and ethically than in how to prompt it well. The lowest means among the verification sub-dimensions belong to fact-checking and identifying hallucinated content. This lines up with literature showing that while many students grasp AI fallibility in theory, they do not regularly follow through in practice

Reliability and validity

Table 3. Reliability and validity indicators

Construct	No. of items	Cronbach's α	Composite reliability (CR)	AVE
GenAI Literacy	16	.89	.90	.58
— Understanding of AI capabilities	4	.82	.83	.55
— Prompt literacy	4	.80	.81	.52
— Awareness of AI limitations	4	.84	.85	.59
— Ethical awareness	4	.81	.82	.53

AI Verification Practices	16	.87	.88	.56
Critical Thinking	12	.85	.86	.54

Because each Cronbach's alpha and composite reliability surpassed the customary .70 cutoff and every AVE cleared .50, the constructs met the requirements for internal consistency and convergent validity. Discriminant validity held up as well: for every construct, the square root of the AVE (.72 to .77) came out higher than its correlations with the remaining two variables (Table 4), satisfying the standard criteria.

Correlation analysis

Table 4. Correlations among study variables

Variable	1	2	3
1. GenAI Literacy	(.76)		
2. AI Verification Practices	.52***	(.75)	
3. Critical Thinking	.46***	.49***	(.74)

Note. Diagonal values in parentheses are the square roots of AVE. *** $p < .001$.

Both AI verification practices ($r = .52$) and critical thinking ($r = .46$) correlate positively and significantly with GenAI literacy, while verification practices show a positive correlation with critical thinking as well ($r = .49$). Ranging from moderate to moderately strong, these correlations align with the proposed mediation model, where literacy connects to critical thinking both on its own and through verification behavior.

Hierarchical regression analysis

Table 5. Hierarchical regression predicting critical thinking

Predictor	Step 1: B (SE)	Step 2: B (SE)
Academic level (postgraduate = 1)	0.18* (0.08)	0.11 (0.07)
Discipline (STEM = 1)	0.09 (0.07)	0.06 (0.06)

Frequency of GenAI use	0.14** (0.05)	0.08 (0.05)
GenAI Literacy		0.41*** (0.05)
Constant	2.98*** (0.14)	1.72*** (0.19)
R ²	.06	.24
ΔR^2	—	.18***
F	8.32***	30.77***

Note. Unstandardized coefficients with standard errors in parentheses; * $p < .05$, ** $p < .01$, *** $p < .001$; N = 400.

The control variables introduced in Step 1 account for only a modest 6% of the variance in critical thinking. Introducing GenAI literacy in Step 2 lifts the explained variance by 18 percentage points ($\Delta R^2 = .18$, $p < .001$), with literacy proving to be a strong, statistically significant positive predictor of critical thinking ($B = 0.41$, $p < .001$); this confirms the hypothesis that it predicts critical thinking in a meaningful way beyond demographic and usage controls.

Mediation analysis

Table 6. Mediation of the GenAI literacy–critical thinking relationship by AI verification practices

Path	Coefficient (B)	SE	95% CI (bootstrap)	p
a: GenAI Literacy → AI Verification Practices	0.56	0.06	[0.44, 0.68]	< .001
b: AI Verification Practices → Critical Thinking (controlling for literacy)	0.29	0.05	[0.19, 0.39]	< .001
c: Total effect (GenAI Literacy → Critical Thinking)	0.41	0.05	[0.31, 0.51]	< .001

c': Direct effect (GenAI Literacy → Critical Thinking, controlling for verification)	0.25	0.05	[0.15, 0.35]	< .001
a × b: Indirect effect	0.16	0.03	[0.10, 0.23]	—

Both path a (GenAI literacy to AI verification practices) and path b (verification practices to critical thinking) are positive and reach statistical significance. Because the bootstrapped 95% confidence interval for the indirect effect ($a \times b = 0.16$, CI [0.10, 0.23]) excludes zero, that indirect effect is statistically significant. Since the direct effect ($c' = 0.25$) stays statistically significant even when controlling for the mediator, this points to partial mediation: AI verification practices account for some of the link between GenAI literacy and critical thinking, but not the whole thing. Roughly 39% of the total effect comes through this indirect pathway ($0.16 / 0.41$), showing that verification practices form a major channel between literacy and critical thinking, though not the only one.

Summary of Hypothesis Testing

Table 7. Summary of hypotheses

Hypothesis	Result
H1: GenAI literacy is positively associated with critical thinking	Supported
H2: GenAI literacy is positively associated with AI verification practices	Supported
H3: AI verification practices are positively associated with critical thinking	Supported
H4: AI verification practices mediate the GenAI literacy–critical thinking relationship	Supported (partial mediation)

DISCUSSION, CONCLUSION, AND RECOMMENDATIONS

These illustrative findings offer a consistent picture of how university students' generative AI literacy connects to their critical thinking: literacy counts, yet much of that value comes from the active step of checking AI-generated material instead of from literacy by itself. The significant positive direct path from GenAI literacy to critical thinking fits with intervention studies showing that deliberate, structured work with tools such as ChatGPT can build stronger reasoning and argumentation (de la Puente, Torres, Troncoso, Meza, & Carrascal, 2024; Deng, Jiang, Yu, Lu, & Liu, 2025). That said, AI verification practices carry a sizable indirect effect, making up an illustrative 39% of the total effect, which ties this hopeful data back to warnings in the literature about students relying too heavily and uncritically on AI (Darwin et al., 2024). The implication is that critical thinking does not grow out of literacy on its own; instead, the gains depend largely on whether students go through the trouble of fact-checking, cross-checking, and vetting AI

text for errors or fabrications, which reflects a growing focus in studies on AI hallucinations and user verification (Alkaissi & McFarlane, 2023; Farquhar, Kossen, Kuhn, & Gal, 2024).

There is a credible cognitive basis for how this works. Merely being aware that generative AI can make mistakes (one part of literacy measured under "awareness of limitations") never forces a student to do hard cognitive work on what the tool produces. The evaluative and inferential skills behind critical thinking are actively put to work by the behavioral habit of verification, whether that means looking for confirming or disconfirming evidence, checking AI claims against course materials, or running code and tracking down citations. Earlier findings point in the same direction: what sets more critical AI users apart from more passive ones is their mode of engagement with AI-generated content, not merely their stated literacy (Chen, Tallant, & Selig, 2025). Because the mediation is only partial, GenAI literacy clearly does some direct work on its own rather than operating entirely through verification habits; for instance, a student with a solid conceptual grasp of AI might anticipate high-risk contexts before they ever start fact-checking.

To sum up, this paper set out a structured quantitative design meant to determine whether verification habits mediate between generative AI literacy and critical thinking across university students in Islamabad, Pakistan. These illustrative findings point toward partial mediation, meaning literacy and verification routines feed into critical thinking both on their own and in tandem within an academic setting full of AI tools. Once empirical data are in hand, the study will face several clear constraints: a cross-sectional approach cannot prove which way causality runs, relying on self-reports instead of direct performance tests introduces known measurement biases (Jin, Martinez-Maldonado, Gašević, & Yan, 2025), and drawing the entire sample from Islamabad limits how far these findings apply to other kinds of institutions or areas of Pakistan.

This framework points to several practical recommendations. First, universities need to teach AI-verification skills as an independent subject. Broad AI literacy cannot simply be counted on to build good checking habits, so course curricula ought to incorporate targeted practice in spotting hallucinated text, fact-checking system outputs, and cross-checking machine claims against reliable source material. Second, because prompt literacy scored comparatively low in the illustrative descriptive data, programs focused on AI literacy might find good reason to concentrate on effective, critical prompting. Third, rules governing academic integrity and AI adoption need to run alongside direct training in verification. Simply understanding the ethics does not mean a student will actually go and verify whether an AI-generated statement is true. Finally, because the mediation shown here was only partial, future campus programs cannot rely on either piece in isolation. Safeguarding and developing students' critical thinking in an era of generative AI means addressing broad AI literacy alongside specific verification practices.

REFERENCES

- Al-Abdullatif, A. M., & Alsubaie, M. A. (2024). ChatGPT in learning: Assessing students' use intentions through the lens of perceived value and the influence of AI literacy. *Behavioral Sciences*, 14(9), Article 845. <https://doi.org/10.3390/bs14090845>
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), Article e35179. <https://doi.org/10.7759/cureus.35179>
- Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, 6, Article 100173. <https://doi.org/10.1016/j.caeo.2024.100173>
- Annapureddy, R., Fornaroli, A., & Gatica-Perez, D. (2025). Generative AI literacy: Twelve defining competencies. *Digital Government: Research and Practice*, 6. <https://doi.org/10.1145/3685680>

- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93–114). Routledge.
- Batista, J., Mesquita, A., & Carnaz, G. (2024). Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. *Information*, 15(11), Article 676. <https://doi.org/10.3390/info15110676>
- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.
- Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, 86(3), 257–295. <https://doi.org/10.1177/23294906231176517>
- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS – Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1(2), Article 100014. <https://doi.org/10.1016/j.chbah.2023.100014>
- Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138, Article 107468. <https://doi.org/10.1016/j.chb.2022.107468>
- Chen, K., Tallant, A. C., & Selig, I. (2025). Exploring generative AI literacy in higher education: Student adoption, interaction, evaluation and ethical perceptions. *Information and Learning Science*, 126(1–2), 132–148. <https://doi.org/10.1108/ILS-10-2023-0160>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.
- Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, 11(1), Article 2290342. <https://doi.org/10.1080/2331186X.2023.2290342>
- de la Puente, M., Torres, J., Troncoso, A. L. B., Meza, Y. Y. H., & Carrascal, J. X. M. (2024). Investigating the use of ChatGPT as a tool for enhancing critical thinking and argumentation skills in international relations debates among undergraduate students. *Smart Learning Environments*, 11, Article 55. <https://doi.org/10.1186/s40561-024-00347-0>
- 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>
- DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Sage.
- Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, 630, 625–630. <https://doi.org/10.1038/s41586-024-07421-0>

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, 9, Article 100436. <https://doi.org/10.1016/j.caeai.2025.100436>

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614. <https://doi.org/10.1146/annurev.psych.58.110405.085542>

Pallant, J. (2010). *SPSS survival manual* (4th ed.). McGraw-Hill.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/BRM.40.3.879>

Sekaran, U., & Bougie, R. (2010). *Research methods for business: A skill building approach* (5th ed.). Wiley.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson.

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206. <https://doi.org/10.1086/651257>