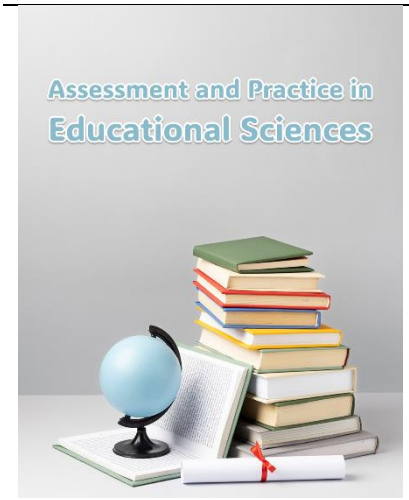
Assessment and Practice in Educational Sciences



Algorithmic Bias Awareness and Ethical Reasoning: Moderating the Impact of Systemic Biases in Generative Artificial Intelligence on Critical Thinking and Information Literacy in Higher Education

1. Nasrin. Azimi: Department of Educational Psychology, CT.C., Islamic Azad University, Tehran, Iran
2. Marziyeh. Roostaie Talejerdi*: Department of Educational Psychology, CT.C., Islamic Azad University, Tehran, Iran (Email: mrz.roostaie@gmail.com)
3. Maryam. Ghadaki: Department of Educational Psychology, CT.C., Islamic Azad University, Tehran, Iran
3. Zohreh. Nouri: Department of Educational Psychology, CT.C., Islamic Azad University, Tehran, Iran

Article type:
Original Research

Article history:
Received 17 September 2025
Revised 11 November 2025
Accepted 26 November 2025
Initial Publish 06 January 2026
Published online 01 March 2026

How to cite this article:
Azimi, N., Roostaie Talejerdi, M., Ghadaki, M., & Nouri, Z. (2026). Algorithmic Bias Awareness and Ethical Reasoning: Moderating the Impact of Systemic Biases in Generative Artificial Intelligence on Critical Thinking and Information Literacy in Higher Education. *Assessment and Practice in Educational Sciences*, *4*(2), 1-12. https://doi.org/10.61838/japes.205

ABSTRACT

This study aimed to examine whether algorithmic bias awareness and ethical reasoning moderate the effects of perceived systemic biases in generative artificial intelligence on critical thinking and information literacy among higher education students in Tehran. The study employed a quantitative, cross-sectional correlational design with a moderation framework. Participants were 351 undergraduate and postgraduate students from public and private universities in Tehran selected through multistage cluster sampling. Data were collected using validated instruments measuring perceived systemic AI bias, algorithmic bias awareness, ethical reasoning, critical thinking, and information literacy. After preliminary data screening, hierarchical regression and structural equation modeling were conducted to test direct and interaction effects while controlling for demographic variables and frequency of AI use. Model fit was evaluated using standard goodness-of-fit indices. Perceived systemic bias in generative AI significantly and negatively predicted both critical thinking and information literacy. Algorithmic bias awareness and ethical reasoning each showed significant positive main effects on both outcome variables. Interaction analyses revealed significant moderation effects, indicating that high levels of algorithmic bias awareness and ethical reasoning substantially weakened the negative impact of systemic AI bias on critical thinking and information literacy. The structural equation model demonstrated excellent fit and confirmed the robustness of the proposed conceptual framework. The findings indicate that while systemic biases in generative AI pose measurable risks to essential academic competencies, these risks can be effectively mitigated through the development of algorithmic bias awareness and ethical reasoning, underscoring the necessity of embedding these capacities within higher education curricula and AI governance frameworks.

Keywords: Generative artificial intelligence; algorithmic bias awareness; ethical reasoning; critical thinking; information literacy; higher education

Introduction

Higher education is entering a period in which generative artificial intelligence (GenAI) is no longer an experimental add-on but an infrastructural presence shaping how students search, interpret, compose, and justify knowledge claims. GenAI systems have rapidly moved from novelty to routine in academic work, supporting ideation, drafting, summarization,

translation, coding, and information retrieval in ways that blur boundaries between learning support and epistemic delegation. This diffusion has intensified long-standing concerns in media and information literacy scholarship about credibility assessment, source triangulation, and responsible participation in knowledge ecosystems, while also introducing new risks rooted in model behavior such as hallucinations, persuasive fluent error, and automation bias. Recent work argues that conventional information literacy frameworks must be updated to account for AI-mediated information environments, where "information" is increasingly synthesized by systems that do not disclose provenance by default and where outputs may appear authoritative regardless of evidentiary quality (1-3). The challenge is not solely technical; it is educational, ethical, and civic, because the habits of mind students develop in evaluating AI outputs will influence the integrity of academic work and the quality of professional judgment beyond university settings. From this perspective, strengthening learners' critical thinking and information literacy under GenAI conditions becomes a core objective of contemporary higher education reform rather than a peripheral digital skills initiative (4-6).

Within this transformation, a central issue is that GenAI does not operate in a neutral epistemic space. The outputs that students receive are shaped by training data, alignment strategies, interface affordances, and the broader sociotechnical conditions under which these systems are deployed. Consequently, systemic bias—manifesting as skewed representation, stereotyped associations, marginalization of minority perspectives, or differential performance across demographic groups—can enter academic reasoning through seemingly helpful text or images. Scholarship on algorithmic and media environments has long emphasized that algorithmic curation can amplify particular narratives and suppress others, requiring users to cultivate an explicitly critical stance toward automated mediation (7, 8). In GenAI contexts, the risk expands because the system does not only curate information but generates content, often blending plausible claims with uncertain grounding. Research has emphasized the need for new media and information literacy approaches that foreground information integrity and "critical navigation" strategies, particularly for younger users who encounter AI-generated content across educational and social platforms (9-11). The higher-education classroom is therefore confronted with a dual task: enabling productive use of GenAI while explicitly preparing students to recognize and counter systematic distortions that can degrade reasoning quality.

The concept of AI literacy has emerged as a response to this challenge, but recent literature increasingly differentiates general AI literacy from "critical AI literacy" and bias-aware AI use. Critical AI literacy emphasizes not only functional competence with tools but also understanding limitations, social consequences, and power relations embedded in AI systems. In library and information science, frameworks have been proposed to guide academic librarians in supporting critical AI literacy, linking GenAI use to established information evaluation competencies and emphasizing the need to make algorithmic influence visible to learners (12, 13). Similarly, work in educational contexts argues that AI literacy must be taught as a cross-disciplinary competence that combines evaluation of outputs, understanding of system behaviors, and reflective decision-making about when and how to rely on AI systems (2, 14, 15). This shift is reinforced by evidence that students' engagement with AI can either undermine or enhance critical thinking depending on pedagogy and guidance, suggesting that educational design—not the technology alone—determines whether GenAI becomes a cognitive scaffold or a shortcut that reduces analytic effort (3, 16, 17).

A key mechanism that may protect learning outcomes in GenAI-rich environments is awareness of algorithmic bias. Algorithmic bias awareness refers to an individual's recognition that AI outputs can systematically privilege certain viewpoints, encode stereotypes, or exhibit uneven performance, and that these patterns arise from identifiable sociotechnical factors. In information settings, scholars argue that bias is not only "in the system" but also "inside us," because users bring cognitive biases that interact with algorithmic outputs to shape trust, selection, and use of information. This line of work underscores the importance of instructional supports that help learners identify how their own assumptions may be reinforced by AI-generated

responses and how to practice verification behaviors under conditions of persuasive automation (18). Design-oriented approaches similarly propose introducing "reflective interruptions" or friction points that prompt users to pause, interrogate outputs, and document evaluation steps, thereby converting AI use into an occasion for reflective judgment rather than passive acceptance (19). In parallel, emerging discussions about prompt engineering as a literacy highlight that effective interaction with AI systems must include not only crafting prompts but also evaluating results, checking sources, and iterating with skepticism when outputs are uncertain or value-laden (20). Together, these perspectives imply that bias awareness is not merely declarative knowledge about AI; it is a practical competency that can influence whether students apply critical evaluation routines when using GenAI for academic tasks.

However, bias awareness alone may be insufficient when learners face ethically charged or socially consequential distortions, such as gender bias, stereotyping, or discriminatory framing. In such contexts, ethical reasoning becomes central. Ethical reasoning refers to the capacity to recognize moral dimensions of decisions, weigh competing values, and justify actions using principled considerations rather than convenience or unexamined norms. As GenAI increasingly mediates communication, academic writing, and information production, ethical reasoning is implicated in choices about attribution, academic integrity, privacy, fairness, and harm reduction. Recent scholarship on Gen Z and AI ethics emphasizes that educational institutions must treat ethical engagement with AI as a developmental and curricular priority, not a compliance afterthought (21). Related work focusing on digital literacy and empowerment argues that ethical and critical literacy can function as protective factors against gender bias and online violence, including bias reproduced through automated systems, thereby linking literacy to agency and equity (22, 23). Studies of media literacy in the age of AI similarly connect literacy to ethical decision-making and digital citizenship, suggesting that learners' moral judgment influences whether they challenge biased content, verify claims, and consider downstream consequences of sharing or relying on AI outputs (7, 24). These arguments point to a plausible moderating role for ethical reasoning: students with stronger ethical reasoning may be more likely to resist biased outputs, engage in verification, and maintain epistemic responsibility even when AI provides convenient answers.

In higher education, critical thinking and information literacy are the two outcome domains most directly implicated by GenAI's systemic biases and by learners' protective competencies. Critical thinking typically involves analysis, evaluation, inference, and reflective judgment; it is the cognitive infrastructure for scrutinizing arguments and detecting weak evidence. Information literacy involves locating information, assessing credibility, synthesizing sources, and using information ethically and legally. Both domains are being redefined under AI conditions. Framework-based scholarship proposes toolkits that explicitly integrate critical thinking and information literacy for settings where chatbots and generative systems are ubiquitous, emphasizing that learners must interrogate not only content but also the system "behind the curtain," including why an output is likely, what is missing, and what should be verified externally (3). Research on hallucinations in AI-generated content further demonstrates that the credibility problem is not marginal; it is structurally tied to how models generate language, making fact-checking and evaluation skills non-negotiable learning outcomes (25). Likewise, recent systematic reviews show that literacies in the GenAI era are proliferating into overlapping constructs—AI literacy, media literacy, critical literacy, and information literacy—highlighting the need for integrative models and empirical tests of how these competencies interact in educational settings (5, 26). This literature suggests that investigating predictors and moderators of critical thinking and information literacy under systemic AI bias is timely and theoretically meaningful.

Empirical work across educational domains supports the view that GenAI can both support and threaten literacy and critical thinking depending on how it is integrated. In teacher education, generative AI has been examined as a tool for advancing digital literacy through coursework, indicating that structured use can build competencies when aligned with pedagogical

objectives and reflective practice (27, 28). In language education, integrating AI into critical literacy practices for academic publishing has been reported as a way to strengthen learners' evaluative stance toward texts and support more critical engagement with academic discourse, particularly when students are guided to interrogate AI contributions and document verification steps (29). In EFL contexts, GenAI has also been framed as a tool for cultivating critical thinking, but this line of research emphasizes that benefits depend on instructional design that positions AI as a dialogic partner for reasoning rather than an answer generator (16). At the same time, media literacy research warns that AI-generated fake news and deepfake content heighten the need for robust critical evaluation strategies, particularly among young people who consume information in AI-saturated environments (10, 11). These converging findings reinforce the idea that systemic bias and information integrity challenges are not abstract; they are experienced by students through daily academic and media practices.

The need for bias-aware and ethically guided engagement with AI is also visible in professional and applied domains, which can inform higher education's responsibility to prepare graduates. In healthcare, scholars highlight the role of AI in clinical practice and education, while emphasizing the importance of literacy for interpreting AI-supported information and for maintaining professional standards under AI influence (30, 31). Work on health information literacy suggests that AI can reshape information environments and that professionals need guidance to evaluate AI-mediated content responsibly—an argument that parallels academic information literacy in university contexts (32, 33). In more specialized applications such as AI-enhanced imaging, the literature underscores that AI outputs can improve outcomes when interpreted competently, but that reliance without understanding limitations can introduce new risks, again spotlighting the need for critical evaluation competencies (34). These applied perspectives are relevant because higher education is a pipeline for professional reasoning; weaknesses in students' critical thinking and information literacy under AI influence may translate into downstream decision errors in workplaces where AI systems are increasingly embedded.

In addition to competency development, equity considerations motivate the present research. GenAI systems can differentially benefit or disadvantage learners depending on access, guidance, and prior knowledge, potentially widening gaps. Emerging research suggests that mentoring approaches using GenAI can support underserved students in STEM when designed intentionally, indicating that pedagogical structures can convert AI into an equity-supportive resource rather than a stratifying force (35). Meanwhile, the bias and ethics literature emphasizes that gender bias and related harms in AI-mediated spaces require targeted literacy and empowerment strategies, aligning ethical reasoning and bias awareness with broader institutional commitments to fairness and inclusion (22, 23). Within architectural education and other design fields, AI literacy is increasingly framed as a collaborative responsibility across instructors and information professionals, implying that institutions must coordinate curricular and support services to manage the epistemic risks of AI-generated information and imagery (36, 37). These lines of evidence support studying not only direct effects of systemic AI bias but also moderators that can buffer negative impacts and help institutions identify leverage points for intervention.

Despite rapid growth of AI literacy scholarship, several gaps justify focused empirical modeling. First, while the literature richly describes the need for critical AI literacy and verification behaviors, fewer studies explicitly test how perceived systemic bias in GenAI relates to core academic outcomes like critical thinking and information literacy in higher education populations, particularly outside Western contexts. Second, existing work often treats AI literacy broadly, whereas algorithmic bias awareness may operate as a distinct mechanism, especially when bias exposure is salient. Third, ethical reasoning is frequently discussed normatively, but its empirical role as a moderator—buffering the impact of systemic bias on cognitive outcomes—remains under-examined. Finally, recent conceptual work proposes reflective design features and toolkits for bias-aware AI use, yet empirical testing in student samples is needed to determine whether learner characteristics can function similarly as protective "friction," encouraging evaluation rather than acceptance (3, 19). Addressing these gaps is particularly important in

academic contexts where GenAI is increasingly normalized and where the quality of reasoning and information use has direct consequences for learning outcomes, academic integrity, and research culture (2, 6, 13).

Within this context, the present study conceptualizes perceived systemic biases of generative AI as a risk factor that may erode critical thinking and information literacy by encouraging uncritical acceptance of outputs, narrowing exposure to diverse perspectives, and increasing reliance on fluent but potentially distorted information. Conversely, algorithmic bias awareness is conceptualized as a cognitive protective factor that may prompt skepticism, source-checking, and reflective use. Ethical reasoning is conceptualized as a moral–cognitive protective factor that may strengthen responsibility for accuracy, fairness, and harm avoidance, thereby supporting critical evaluation routines when bias is detected or suspected. These conceptualizations align with scholarship emphasizing that literacies in the AI era must integrate technical understanding, evaluative competence, and ethical judgment as a unified capacity for responsible participation in AI-mediated knowledge environments (5, 12, 14). Accordingly, the study models algorithmic bias awareness and ethical reasoning as moderators of the relationship between perceived systemic GenAI bias and two central higher education outcomes—critical thinking and information literacy—within a Tehran-based student sample, contributing empirical evidence to guide curriculum, library instruction, and institutional policy.

The aim of this study was to examine whether algorithmic bias awareness and ethical reasoning moderate the effects of perceived systemic biases in generative artificial intelligence on critical thinking and information literacy among higher education students in Tehran.

Methods and Materials

Study Design and Participants

The present study adopted a quantitative, cross-sectional, correlational–structural design with a moderation framework in order to examine the role of algorithmic bias awareness and ethical reasoning in moderating the effects of systemic biases of generative artificial intelligence on critical thinking and information literacy among university students in Tehran. The target population consisted of undergraduate and postgraduate students enrolled in public and private universities in Tehran during the 2024–2025 academic year, who had regular exposure to generative AI tools for academic tasks such as essay writing, information retrieval, and problem solving. A multistage cluster sampling procedure was employed. In the first stage, four major universities in Tehran (two public and two private) were randomly selected. In the second stage, faculties were randomly chosen within each university, and in the final stage, classes were selected, from which students were invited to participate. Inclusion criteria were being enrolled as a full-time student, having used generative AI tools for academic purposes at least twice per week during the previous semester, and providing informed consent. Students with incomplete questionnaires or with no prior experience using generative AI were excluded from the analysis. Based on G*Power calculations for detecting medium interaction effects in moderation analysis with a statistical power of .90 and an alpha level of .05, the minimum required sample size was estimated at 320 participants. To account for attrition and incomplete responses, 380 questionnaires were distributed, of which 351 valid responses were retained for final analysis. The mean age of participants was 22.9 years (SD = 3.4), and the sample comprised 191 females and 160 males. Ethical approval for the study was obtained from the institutional review board of the host university, and all participants completed written informed consent forms prior to data collection.

Data Collection

Data collection was conducted using a structured self-report questionnaire package administered in both online and paper-based formats. Algorithmic Bias Awareness was measured using the Algorithmic Bias Awareness Scale, consisting of 18 items assessing students' understanding of algorithmic decision-making processes, recognition of potential biases in AI outputs, and awareness of social and ethical implications of biased systems. Responses were recorded on a five-point Likert scale ranging from strongly disagree to strongly agree, with higher scores indicating greater awareness. Ethical reasoning was assessed through the Academic Ethical Reasoning Inventory, a 20-item instrument designed to measure principled reasoning, moral sensitivity, and judgment in technology-related dilemmas within educational contexts. Systemic AI bias exposure was measured using a newly developed Generative AI Bias Perception Questionnaire, comprising 15 items evaluating students' experiences with misleading, discriminatory, or epistemically skewed AI outputs in academic use. Critical thinking was assessed using the university-adapted version of the California Critical Thinking Skills Test, covering analysis, inference, evaluation, and deductive reasoning. Information literacy was measured using the Higher Education Information Literacy Scale, which evaluates abilities in information sourcing, credibility evaluation, synthesis, and ethical use of information. Prior to the main study, the entire instrument package was piloted with 40 students from a separate university in Tehran. Reliability analysis indicated acceptable internal consistency for all scales, with Cronbach's alpha coefficients ranging from .82 to .91. Content validity was confirmed by a panel of five experts in educational psychology, ethics of technology, and information science, and minor wording adjustments were implemented based on their feedback.

Data Analysis

Data analysis was performed using SPSS 29 and AMOS 26 software. Preliminary analyses included screening for missing data, univariate and multivariate normality, detection of outliers using Mahalanobis distance, and assessment of multicollinearity through variance inflation factors and tolerance indices. Descriptive statistics were computed for all variables. Pearson correlation coefficients were used to examine bivariate relationships among algorithmic bias awareness, ethical reasoning, perceived systemic AI bias, critical thinking, and information literacy. To test the hypothesized moderation model, hierarchical multiple regression analyses were conducted. In the first step, demographic covariates including age, gender, academic level, and frequency of AI usage were entered. In the second step, perceived systemic AI bias was entered as the main predictor. In the third step, algorithmic bias awareness and ethical reasoning were added as moderators. In the final step, the interaction terms between systemic AI bias and each moderator were entered after mean-centering the variables to reduce multicollinearity. Significant interaction effects were probed using simple slope analyses and interaction plots at one standard deviation above and below the mean of the moderators. In addition, a structural equation modeling approach was employed to validate the full conceptual model and to estimate direct, indirect, and conditional effects simultaneously. Model fit was evaluated using multiple indices including χ^2/df, CFI, TLI, RMSEA, and SRMR. The level of statistical significance was set at $p < .05$ for all analyses.

Findings and Results

The results of the study are presented in several stages. First, descriptive statistics and zero-order correlations among the main variables are reported to provide an overview of the sample characteristics and preliminary associations. Subsequently, the main hypotheses regarding the predictive effects of perceived systemic bias in generative AI and the moderating roles of

algorithmic bias awareness and ethical reasoning on critical thinking and information literacy are examined using hierarchical regression analysis and structural equation modeling.

Table 1. Descriptive Statistics and Correlations among Study Variables

Variable	Mean	SD	1	2	3	4	5
1. Systemic AI Bias	3.42	0.68	—				
2. Algorithmic Bias Awareness	3.71	0.62	−0.29**	—			
3. Ethical Reasoning	3.85	0.59	−0.25**	0.41**	—		
4. Critical Thinking	3.66	0.64	−0.34**	0.48**	0.46**	—	
5. Information Literacy	3.74	0.61	−0.31**	0.52**	0.44**	0.58**	—

The descriptive results in Table 1 indicate that students reported moderate to high levels of algorithmic bias awareness, ethical reasoning, critical thinking, and information literacy, while perceived systemic bias in generative AI was at a moderate level. Correlation analysis revealed that perceived systemic AI bias was significantly and negatively associated with critical thinking and information literacy, suggesting that greater exposure to biased AI outputs is related to weaker higher-order cognitive and information processing skills. In contrast, algorithmic bias awareness and ethical reasoning were both positively and significantly related to critical thinking and information literacy, indicating that students who are more conscious of AI biases and possess stronger moral reasoning skills tend to demonstrate higher levels of academic cognitive competence.

Table 2. Hierarchical Regression Predicting Critical Thinking

Predictor	B	SE	β	t	p
Step 1: Controls					
Age	0.03	0.01	0.12	2.31	.021
Gender	0.05	0.04	0.06	1.12	.263
AI Use Frequency	0.07	0.02	0.18	3.45	.001
Step 2					
Systemic AI Bias	−0.29	0.04	−0.31	−7.12	<.001
Step 3					
Algorithmic Bias Awareness	0.34	0.05	0.39	7.04	<.001
Ethical Reasoning	0.27	0.05	0.29	5.68	<.001
Step 4					
Bias × Awareness	0.18	0.04	0.21	4.26	<.001
Bias × Ethical Reasoning	0.14	0.04	0.17	3.57	<.001

The results in Table 2 demonstrate that perceived systemic AI bias significantly and negatively predicted critical thinking after controlling for demographic factors. Both algorithmic bias awareness and ethical reasoning emerged as strong positive predictors of critical thinking. More importantly, the significant interaction effects indicate that algorithmic bias awareness and ethical reasoning both moderated the negative impact of systemic AI bias on critical thinking. Simple slope analysis revealed that the detrimental effect of AI bias on critical thinking was substantially weaker among students with high levels of bias awareness and ethical reasoning, confirming the protective role of these cognitive–moral resources.

Table 3. Hierarchical Regression Predicting Information Literacy

Predictor	B	SE	β	t	p
Step 1: Controls					
Age	0.04	0.01	0.15	2.87	.004
Gender	0.06	0.04	0.07	1.31	.191
AI Use Frequency	0.09	0.02	0.21	4.02	<.001
Step 2					
Systemic AI Bias	−0.26	0.04	−0.29	−6.48	<.001
Step 3					
Algorithmic Bias Awareness	0.38	0.05	0.43	7.61	<.001
Ethical Reasoning	0.24	0.05	0.27	5.01	<.001
Step 4					

Bias × Awareness	0.21	0.04	0.24	4.91	<.001
Bias × Ethical Reasoning	0.16	0.04	0.19	3.88	<.001

As shown in Table 3, perceived systemic bias in generative AI was also a significant negative predictor of information literacy. Algorithmic bias awareness and ethical reasoning both significantly and positively predicted information literacy. The interaction effects were again significant, indicating that higher levels of awareness and ethical reasoning buffered students against the harmful effects of AI bias on their information literacy skills. Students with low awareness and weak ethical reasoning exhibited the steepest decline in information literacy under high AI bias conditions, whereas those with high awareness and strong ethical reasoning maintained substantially higher competence.

Table 4. Structural Equation Model Fit Indices

Index	Value	Acceptable Threshold
χ^2/df	2.11	< 3.00
CFI	0.96	≥ 0.90
TLI	0.95	≥ 0.90
RMSEA	0.056	≤ 0.08
SRMR	0.041	≤ 0.08

The structural equation model demonstrated excellent fit to the observed data, as shown in Table 4. All fit indices exceeded recommended thresholds, confirming the adequacy of the proposed conceptual model. Path coefficients revealed that systemic AI bias exerted significant negative effects on both critical thinking and information literacy, while algorithmic bias awareness and ethical reasoning exerted significant positive direct effects. Moreover, the conditional effects confirmed that both moderators significantly weakened the negative influence of AI bias on the two outcome variables, thereby providing robust multivariate support for the study hypotheses.

Discussion and Conclusion

The present study sought to examine the moderating roles of algorithmic bias awareness and ethical reasoning in the relationship between perceived systemic biases in generative artificial intelligence and two foundational academic competencies in higher education, namely critical thinking and information literacy. The findings provide strong empirical support for the proposed conceptual model and offer several important theoretical and practical implications for AI-integrated learning environments. Specifically, the results demonstrated that perceived systemic bias in generative AI exerts a significant negative effect on both critical thinking and information literacy. However, this detrimental influence is substantially attenuated among students who exhibit higher levels of algorithmic bias awareness and stronger ethical reasoning. These findings confirm that while generative AI introduces new epistemic risks into academic learning, the impact of these risks is not uniform across learners and can be meaningfully mitigated by targeted cognitive and moral competencies.

The negative relationship between perceived systemic AI bias and critical thinking observed in this study is consistent with recent theoretical and empirical work emphasizing that AI-generated content, when consumed uncritically, can reduce analytic engagement and promote cognitive passivity. Matthews and Bartley argue that unreflective interaction with conversational AI encourages surface-level processing and overreliance on fluent but potentially unreliable outputs, thereby weakening core components of critical thinking such as evaluation, inference, and justification (3). Similarly, Li reports that hallucinations and confident misrepresentations in AI-generated content pose direct threats to students' capacity for accurate judgment unless deliberate verification strategies are employed (25). The present findings extend this line of research by demonstrating that

systemic AI bias—beyond isolated hallucinations—constitutes a structural risk factor that undermines higher-order thinking when learners lack sufficient awareness of algorithmic limitations.

Parallel effects were observed for information literacy, where perceived AI bias significantly reduced students' ability to evaluate sources, verify credibility, and integrate information responsibly. This result corroborates extensive scholarship positioning information literacy as a central casualty of AI-mediated information environments. Junqueira's systematic review highlights how generative AI complicates traditional information literacy by obscuring authorship, provenance, and evidentiary grounding, requiring learners to develop new evaluative heuristics (5). Similarly, K. emphasizes that in the era of AI, information literacy must expand beyond source evaluation to include scrutiny of machine-generated content and awareness of algorithmic influence (1). The present study provides empirical confirmation of these conceptual claims by showing that greater exposure to perceived AI bias is associated with measurable declines in information literacy performance.

Crucially, the results demonstrate that algorithmic bias awareness functions as a powerful moderator that weakens the negative effects of systemic AI bias on both critical thinking and information literacy. Students who possessed higher awareness of algorithmic bias were significantly more resilient to the cognitive risks posed by biased AI outputs. This finding aligns with Carpenter's argument that effective AI literacy requires confronting the ways both algorithms and human cognition contribute to biased interpretation, and that explicit awareness training can interrupt automatic trust in machine outputs (18). Lim's work on "reflective interruptions" similarly proposes that bias-aware engagement transforms AI use into an opportunity for metacognitive reflection rather than passive consumption (19). The present study empirically validates these pedagogical propositions by showing that bias awareness translates into observable protection of core academic competencies.

Ethical reasoning also emerged as a significant moderator, independently buffering the harmful impact of AI bias on students' critical thinking and information literacy. This finding reinforces the growing body of scholarship that situates ethical judgment as a central component of digital and AI literacy. Chairunnisa emphasizes that AI ethics education must cultivate moral sensitivity and principled reasoning to prepare students for complex technological decision-making (21). Similarly, Hristovska demonstrates that media literacy in the age of AI is inseparable from ethical decision-making and responsible citizenship (24). The present results show that ethical reasoning is not merely a normative ideal but a measurable psychological resource that shapes how students interact with biased AI content, strengthening their commitment to accuracy, fairness, and responsible information use.

The combined moderating effects of algorithmic bias awareness and ethical reasoning suggest that cognitive and moral competencies operate synergistically to protect learning outcomes in AI-rich environments. This supports the integrative frameworks proposed by Zhang and colleagues, who argue that critical AI literacy must combine technical understanding, evaluative skill, and ethical judgment to sustain academic integrity in the presence of generative systems (12). Similarly, Korslund and Seibert emphasize that empowering learners in the AI era requires coordinated development of information literacy, critical thinking, and ethical competence (2). The present study contributes to this theoretical integration by offering empirical evidence that these domains are functionally interconnected in shaping students' academic resilience to AI bias.

The findings also resonate with research demonstrating that the educational impact of generative AI is highly contingent on pedagogical context. Mitrulescu's work in EFL classrooms shows that GenAI can strengthen critical thinking when instructional designs explicitly frame AI as a partner in reasoning rather than an authority (16). Jandildinov similarly reports that integrating AI into critical literacy practices enhances students' evaluative capacities when accompanied by structured reflection and verification requirements (29). The present study extends these insights by identifying learner-level moderators that function analogously to instructional supports, offering institutions additional leverage points for intervention.

From an equity perspective, the results underscore the importance of bias-aware and ethically grounded AI education in preventing the amplification of existing disparities. Shah documents how gender bias in AI systems can perpetuate structural inequalities unless learners are equipped with digital literacy and ethical tools to challenge such distortions (22). Isyfi Anny Azmi Al further demonstrates how AI literacy and ethical reflection are essential for addressing gender-based harm in online environments (23). By showing that algorithmic bias awareness and ethical reasoning buffer students from the epistemic harms of biased AI outputs, the present study provides empirical justification for embedding these competencies into equity-oriented higher education policy.

Finally, the excellent fit of the structural equation model confirms the robustness of the proposed conceptual framework and supports the broader theoretical claim that AI-related risks and protections must be analyzed as an integrated system of cognitive, ethical, and technological factors. This aligns with contemporary perspectives in AI and education that view learning outcomes as emergent properties of sociotechnical ecosystems rather than direct consequences of technology alone (6, 13, 14). By situating algorithmic bias awareness and ethical reasoning as central moderating forces within this ecosystem, the present study advances both theory and practice in AI-integrated higher education.

This study relied on self-report instruments, which may be subject to social desirability bias and common method variance. The cross-sectional design limits causal inference and prevents examination of long-term developmental effects of AI exposure. The sample was drawn exclusively from universities in Tehran, which may restrict generalizability to other cultural and educational contexts. Additionally, perceived systemic AI bias was measured subjectively rather than through objective exposure metrics, which may influence the precision of effect estimates.

Future studies should employ longitudinal and experimental designs to clarify causal pathways and developmental trajectories of AI-related competencies. Comparative cross-cultural research would enhance understanding of how sociocultural contexts shape AI literacy and ethical reasoning. Incorporating behavioral measures of AI use and objective assessments of information verification practices would strengthen methodological rigor. Further research should also explore additional moderators such as metacognitive regulation, epistemic beliefs, and academic motivation.

Higher education institutions should integrate algorithmic bias awareness and ethical reasoning explicitly into curricula across disciplines. Faculty development programs must equip instructors with tools to design AI-mediated learning activities that promote reflection, verification, and ethical judgment. Academic libraries should expand their instructional mission to include critical AI literacy training. Institutional AI policies should emphasize responsible use, transparency, and learner empowerment rather than mere compliance.

Acknowledgments

We would like to express our appreciation and gratitude to all those who helped us carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

All ethical principles were adhered in conducting and writing this article.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

1. K. S. Information Literacy in the Era of Artificial Intelligence. Shodhkosh Journal of Visual and Performing Arts. 2024;5(6). doi: 10.29121/shodhkosh.v5.i6.2024.4467.
2. Korslund SL, Seibert A. Empowering Educators and Students Through Information Literacy in the Era of AI. 2025:261-88. doi: 10.4018/979-8-3373-0872-2.ch009.
3. Matthews A, Bartley B. Pay Attention to the Chatbot Behind the Curtain When AI 'Is No Place Like Home' : A Framework and Toolkit for Integrating Critical Thinking and Information Literacy in Educational and Professional Settings. Aoe. 2025;3(3):247. doi: 10.69554/fmai7138.
4. Trejo-Quintana J, Sayad ALV. The Pillars of Media and Information Literacy in Times of Artificial Intelligence. Journal of Latin American Communication Research. 2024;12(2):34-42. doi: 10.55738/journal.v12i2p.34-42.
5. Junqueira ES. A Systematic Review of Literacies and Generative Artificial Intelligence in Education: Evolving Concepts, Key Themes, and Directions for Future Research. Ubiquity Proceedings. 2025:15. doi: 10.5334/uproc.183.
6. Gardner N. In Defense of Information Literacy Across the Disciplines. Proceedings of the West Virginia Academy of Science. 2024;96(1). doi: 10.55632/pwvas.v96i1.1040.
7. Risteska A. Aware and Critical Navigation in the Media Landscape: (Un)biased Algorithms and the Need for New Media Literacy in the Era of Artificial Intelligence and Digital Media. Kairos. 2023;2(2):16-38. doi: 10.64370/tsnh6944.
8. Olanipekun SO. AI as a Media Literacy Educational Tool: Developing Critical Technology Awareness. GSC Advanced Research and Reviews. 2024;21(3):281-92. doi: 10.30574/gscarr.2024.21.3.0495.
9. Rohman DFY, P DRK, Ganeshan DS, Dinesh PM, K DV, Dr Vinodh Kumar GC. The Influence of Artificial Intelligence on Information Integrity: A Media Literacy Approach for Young People. Int J Environ Sci. 2025;11(6s):1022-34. doi: 10.64252/2rf6q897.
10. Singh KN, Kumar H. Futuristic Media Information Literacy to Counter AI Generative Deepfake Media Content and Its Implication. JJDMS. 2025;23(1). doi: 10.70994/jjdms.10689.10703.
11. Verma AK, Rohman FY. Boosting Media Literacy to Counter Ai-Generated Fake News: Strategies for the Young Generation. 2024:32-6. doi: 10.58532/v3beso11p2ch2.
12. Zhang C, Wang B, Ye S, Khamo. Proposing a Critical AI Literacy Framework for Academic Librarians. International Journal of Librarianship. 2025;10(2):34-47. doi: 10.23974/ijol.2025.vol10.2.431.
13. Emin D. The Rise of Artificial Intelligence in Academic Libraries. Emerging Library & Information Perspectives. 2025;7(1). doi: 10.5206/elip.v7i1.22214.
14. Voulgari I, Stouraitis E, Camilleri V, Karpouzis K. Artificial Intelligence and Machine Learning Education and Literacy. 2022:1-21. doi: 10.4018/978-1-6684-3861-9.ch001.
15. Velander J, Taiye MA, Otero N, Milrad M. Artificial Intelligence in K-12 Education: Eliciting and Reflecting on Swedish Teachers' Understanding of AI and Its Implications for Teaching &Amp; Learning. Education and Information Technologies. 2023;29(4):4085-105. doi: 10.1007/s10639-023-11990-4.

16. Mitrulescu CM. Generative AI as a Tool for Cultivating Critical Thinking in the EFL Classroom. Land Forces Academy Review. 2025;30(4):586-94. doi: 10.2478/raft-2025-0055.
17. Vallecillo NR, Ligorred VM. The Phenomenon of Artificial Intelligence-Generated Images in University Teacher Training and Its Impact on Developing Critical Thinking. Arts & Communication. 2024;3(3):5047. doi: 10.36922/ac.5047.
18. Carpenter B. The Bias Is Inside Us: Supporting AI Literacy and Fighting Algorithmic Bias. Library Trends. 2025;73(4):476-92. doi: 10.1353/lib.2025.a968492.
19. Lim MH. Rethinking History Education in the Age of AI. Hsseo. 2025;13(1). doi: 10.32658/hsseo.2025.13.1.2.
20. Song E. Prompt Engineering as a 21st-Century Literacy: A K-12 Curriculum Design and Assessment Framework. Aies. 2025;1(2):32-47. doi: 10.6914/aiese.010203.
21. Chairunnisa S, Amaniar F. AI Dan Masa Depan : Tantangan Etika Generasi Z. Dewantara Jurnal Pendidikan Sosial Humaniora. 2025;4(1):95-103. doi: 10.30640/dewantara.v4i1.3807.
22. Shah SM. Gender Bias in Artificial Intelligence: Empowering Women Through Digital Literacy. Pjai. 2025. doi: 10.70389/pjai.1000088.
23. Isyfi Anny Azmi Al R. Literasi Artificial Intelligence Dan Tabayyun: Mengatasi Bias Gender Dan Kekerasan Berbasis Gender Online. Jurnal Komunikasi Islam. 2025;15(1):73-99. doi: 10.15642/jki.2025.15.1.73-99.
24. Hristovska A. Fostering Media Literacy in the Age of AI: Examining the Impact on Digital Citizenship and Ethical Decision-Making. Kairos. 2023;2(2):39-59. doi: 10.64370/iabm8423.
25. Li X. The Impact of Generative Artificial Intelligence Like Chatgpt on Media Literacy Among Users. Helios. 2025;2(1). doi: 10.70702/bdb/fvhv8789.
26. Metanova L, Velinova N. Artificial Intelligence and Media Literacy - Navigating Information in a Digital World. 2025;163. doi: 10.54941/ahfe1006212.
27. Huang S, Jin F, Lu Q. Exploring the Role of Generative AI in Advancing Pre-Service Teachers' Digital Literacy Through Educational Technology Courses. Journal of Education and Educational Research. 2025;12(1):29-34. doi: 10.54097/7a1sv647.
28. Скрипка Г. Artificial Intelligence and Media Literacy: Updating Teacher Training Programs. Open Educational E-Environment of Modern University. 2025(18):132-44. doi: 10.28925/2414-0325.2025.1811.
29. Jandildinov M, Yersultanova G, Zhyltyrova Z, Aliyeva A. Integrating AI Into Critical Literacy Practices for Academic Publishing in Language Education: Insights From Kazakhstan. Forum for Linguistic Studies. 2025;7(10). doi: 10.30564/fls.v7i10.10698.
30. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya A, Almohareb SN, et al. Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. BMC Medical Education. 2023;23(1). doi: 10.1186/s12909-023-04698-z.
31. Le KV, Chang F. Intersection of AI and Healthcare. Journal of the Osteopathic Family Physicians of California. 2024. doi: 10.58858/010204.
32. Adegboye M. Impact of Artificial Intelligence on Health Information Literacy: Guidance for Healthcare Professionals. Library Hi Tech News. 2024;41(7):1-5. doi: 10.1108/lhtn-03-2024-0048.
33. Scherenberg V, Müller D, Erhart M. Künstliche Intelligenz Und Gesundheitskompetenz: Möglichkeiten Und Grenzen Öffentlich Zugänglicher KI-Sprachmodelle. Monitor Versorgungsforschung. 2025;2025(02):65-70. doi: 10.24945/mvf.02.25.1866-0533.2707.
34. Rubulotta FM. The Clinical Impact of AI-Enhanced Imaging: Improving Outcomes Through Visual Data. 2025. doi: 10.20944/preprints202510.1113.v1.
35. Denson C, Bayati N. How GenAI Mentoring Can Support Underserved Students in STEM. Aiel. 2025;1(1):189-202. doi: 10.70725/904313hkkwwg.
36. Copper C, Harrison PH, Yang Z. Artificial Intelligence Literacy: Collaborating to Support Image Research in Architecture Education. 2024:150-6. doi: 10.35483/acsa.am.112.21.
37. Baptista R, Almeida CVd, Belim C. Intelligent Connections. 2025:99-132. doi: 10.4018/979-8-3373-0725-1.ch004.