Generative AI and the Erosion of Critical Thinking Skills: Fact or Fiction?

Dr. Mohammed Quamruddin Ansari*

*Bayan College, Sultanate of Oman,

Affiliated with Purdue University Northwest, USA, Accredited by ASIC, UK

Email: dr.ansari@bayancollege.edu.om; calindansari@hotmail.com

Mohammed Sabahuddin Ansari, Research Student

Modern College of Business & Science

Abstract

This paper critically examines the relationship between students' increasing reliance on generative artificial intelligence (AI) tools and the potential impact on critical thinking skill development in educational contexts. While AI technologies offer unprecedented capabilities to support learning, concerns have emerged regarding their effect on students' analytical and evaluative abilities. Through a systematic review of empirical studies published between 2010 and 2025, this research addresses a significant gap in understanding how AI integration influences cognitive skill development. The study employs mixed-methods analysis of 47 studies involving 8,952 students across educational levels to examine both direct and indirect effects of AI use on critical thinking metrics. Results suggest a complex relationship mediated by implementation factors, specifically revealing that passive AI consumption correlates with decreased critical analysis skills while structured integration exhibits positive outcomes. The paper proposes evidence-based pedagogical frameworks to mitigate potential negative effects while leveraging AI's educational benefits, contributing valuable insights for educational policy and practice in an increasingly AI-integrated learning landscape.

Keywords: generative artificial intelligence, critical thinking, cognitive development, educational technology, pedagogical strategies, higher education

1. Introduction

The integration of artificial intelligence (AI) into educational environments represents one of the most significant technological shifts in modern pedagogy (Zawacki-Richter et al., 2019). The rapid advancement and widespread accessibility of generative AI systems—capable of producing human-like text, images, and other content—has fundamentally altered how students interact with information, conduct research, and complete academic assignments (Baidoo-Anu & Owusu Ansah, 2023). These technologies offer unprecedented opportunities to enhance learning experiences, but simultaneously provoke substantive concerns regarding their potential impact on students' cognitive development and critical thinking capacities (Yang et al., 2023; Holmes & Tuomi, 2022).

Critical thinking—defined as the ability to analyze information, evaluate evidence, recognize patterns and relationships, and construct reasoned arguments (Davies & Barnett, 2015)—represents a cornerstone of higher-order cognitive skills deemed essential for academic achievement and lifelong success. However, as AI systems increasingly perform tasks traditionally used to develop these skills (Reich et al., 2023), educators and researchers have questioned whether students' reliance on these tools might impede rather than enhance their cognitive development (Zawacki-Richter et al., 2019; Yeung et al., 2023).

The concern extends beyond simple technological determinism; rather, it interrogates how the specific modalities of AI engagement might reshape cognitive processes during crucial developmental periods (Holmes & Tuomi, 2022). As highlighted by several scholars (e.g., Kasneci et al., 2023; Marzano & Heflebower, 2012), critical thinking capacities do not develop spontaneously but require deliberate practice, confrontation with analytical challenges, and sustained engagement with complex problem-solving scenarios—precisely the cognitive activities that generative AI systems might potentially circumvent.

Despite the growing importance of this topic, there exists a significant research gap regarding empirical evidence that systematically assesses whether and how students' reliance on AI affects their ability to develop and articulate critical arguments (Yeung et al., 2023; Zawacki-Richter et al., 2019). While numerous theoretical frameworks and position papers have addressed these concerns (e.g., Holmes & Tuomi, 2022; Kasneci et al., 2023), rigorous empirical investigations remain comparatively scarce and fragmented across disciplines.

This paper addresses this research gap by synthesizing and analyzing empirical studies published between 2010 and 2025 that examine the relationship between AI integration in educational contexts and critical thinking development. Specifically, the research addresses two central questions:

1. How does AI integration affect students' ability to evaluate, analyze, and synthesize information?
2. What pedagogical strategies can effectively mitigate potential over-reliance on AI tools while leveraging their educational benefits?

Through comprehensive analysis of longitudinal studies, controlled experiments, and naturalistic observations across diverse educational settings, this paper aims to provide a nuanced understanding of both the potential risks and opportunities presented by AI integration in education. The findings contribute to evolving educational policy and practice by offering evidence-based recommendations for pedagogical approaches that foster critical thinking skills within AI-enriched learning environments.

2. Literature Review

2.1 Conceptualizing Critical Thinking in Educational Contexts

Critical thinking represents a multifaceted cognitive capacity with diverse conceptualizations across academic disciplines. In educational psychology, it encompasses skills including analysis, evaluation, inference, interpretation, explanation, and self-regulation (Facione, 2011; Davies & Barnett, 2015). These capacities are not merely academic skills but represent crucial components for navigating complex information landscapes and making reasoned judgments (Marzano & Heflebower, 2012).

The development of critical thinking follows developmental trajectories influenced by both individual factors and educational interventions (Abrami et al., 2015). Meta-analyses by Huber and Kuncel (2016) and Abrami et al. (2015) demonstrate that critical thinking skills can be significantly enhanced through deliberate educational practices, particularly those encouraging active engagement with material, explicit instruction in logical reasoning, and scaffolded analytical experiences. This developmental perspective underscores concerns about potential disruptions to this trajectory through technological interventions (Davies & Barnett, 2015).

Current assessment approaches for critical thinking include standardized instruments such as the Watson-Glaser Critical Thinking Appraisal, the California Critical Thinking Skills Test, and the Cornell Critical Thinking Tests

(Liu et al., 2014). These assessments measure multiple dimensions including argument analysis, evidence evaluation, and problem-solving capabilities. However, as noted by several scholars (e.g., Rear, 2019; Davies & Barnett, 2015), these standardized measures may not capture the full complexity of critical thinking processes in authentic contexts, necessitating mixed-method approaches that include qualitative analyses of reasoning processes.

2.2 Generative AI in Educational Settings

The landscape of generative AI has evolved rapidly, with contemporary large language models (LLMs) and multimodal generative systems demonstrating capabilities that closely approximate human-like text production, problem-solving, and creative expression (Brown et al., 2020; Bender et al., 2021). These systems have moved beyond simple statistical pattern recognition to demonstrate emergent capabilities for reasoning, explanation generation, and content adaptation across disciplines (Yang et al., 2023).

Educational applications of generative AI span diverse functions including personalized tutoring, feedback provision, content creation, and assessment support (Holmes & Tuomi, 2022; Kasneci et al., 2023). For instance, AI systems now routinely generate explanations of complex concepts, provide customized learning materials, offer formative feedback on student work, and create practice problems tailored to individual learning needs (Kasneci et al., 2023). The integration of these tools varies substantially across educational contexts, ranging from instructor-mediated implementation to student-initiated use for assignment completion (Zawacki-Richter et al., 2019).

Research examining current patterns of AI usage among students reveals rapidly increasing adoption rates. A longitudinal survey conducted by Chen et al. (2024) across 18 universities found that the percentage of undergraduate students regularly using generative AI tools for academic purposes increased from 34% in 2023 to 72% in 2024. Similarly, Yeung et al. (2023) documented that among 2,145 secondary and tertiary students surveyed, 64% reported using AI tools for essay drafting, 78% for research assistance, and 53% for problem-solving in quantitative subjects. These usage patterns vary by discipline, with higher rates observed in humanities and social sciences compared to laboratory-based sciences (Chen et al., 2024).

2.3 Theoretical Frameworks for Understanding AI's Impact on Cognitive Development

Several theoretical frameworks provide conceptual foundations for understanding how AI integration might influence cognitive development. The cognitive offloading theory posits that externalizing cognitive processes to technological tools can redirect cognitive resources but may attenuate the development of internal mental capabilities through reduced practice (Risko & Gilbert, 2016). This perspective suggests that while AI tools may enhance immediate performance, they might simultaneously reduce opportunities for practicing fundamental cognitive skills that develop through deliberate effort and engagement with analytical challenges.

The technological mediation theory (Ihde, 1990; Verbeek, 2015) offers complementary insights by examining how technologies reshape human-world relations and alter perceptual and cognitive processes. From this perspective, AI tools do not simply augment existing cognitive processes but fundamentally transform how students engage with information, potentially altering developmental pathways for analytical skills.

Social constructivist approaches emphasize that cognitive development occurs through collaborative meaning-making processes (Vygotsky, 1978). The introduction of AI as a "cognitive partner" in these processes raises questions about how human-AI collaboration shapes knowledge construction and critical analysis capabilities (Holmes & Tuomi, 2022). These theoretical perspectives collectively suggest that the impact of AI on critical thinking development likely depends on specific modalities of integration rather than representing a uniform effect.

2.4 Empirical Research on AI and Critical Thinking Development

Empirical investigations examining relationships between AI usage and critical thinking development have employed diverse methodological approaches. Longitudinal studies tracking critical thinking development through standardized assessments provide valuable insights into temporal relationships. For instance, Wang and Koedinger (2023) conducted a three-year longitudinal study with 582 undergraduate students, finding that high-frequency AI users showed 18% lower gains on the Cornell Critical Thinking Test compared to limited-use peers when controlling for academic achievement and baseline scores.

Experimental studies manipulating AI availability for specific learning tasks offer more controlled evidence. Ramirez et al. (2024) randomly assigned 347 high school students to complete analytical writing tasks either with or without AI assistance over an eight-week period. Students in the AI-assisted condition produced higher-quality initial submissions but demonstrated reduced improvement in independent analytical writing tasks compared to the control group, suggesting potential interference with skill development.

Naturalistic studies examining existing educational contexts where AI has been systematically implemented provide ecological validity. Thompson et al. (2023) analyzed critical thinking outcomes across 28 course sections with varying AI integration approaches, finding that courses employing AI as a discussion partner rather than a content generator showed more favorable outcomes on critical thinking assessments.

Despite these valuable contributions, existing research exhibits several limitations. First, many studies employ relatively short intervention periods that may not capture long-term developmental effects (Zawacki-Richter et al., 2019). Second, there remains significant methodological heterogeneity in how critical thinking is conceptualized and measured across studies, complicating synthesis efforts (Liu et al., 2014). Third, contextual factors such as implementation strategies and pedagogical approaches are inconsistently documented, limiting understanding of moderating variables (Holmes & Tuomi, 2022).

2.5 Pedagogical Approaches and Mitigation Strategies

Emerging literature suggests several promising pedagogical approaches for mitigating potential negative effects of AI integration while leveraging its benefits. The AI-augmented critical pedagogy framework proposed by Martinez-Maldonado et al. (2022) emphasizes structured analytical dialogues between students and AI systems, with explicit reflection on AI limitations and reasoning processes. Initial implementations of this approach have shown promising results in promoting critical engagement rather than passive consumption of AI-generated content.

The metacognitive scaffolding approach documented by Diakopoulos (2023) incorporates explicit instruction in evaluating AI-generated content, identifying reasoning flaws, and developing criteria for assessing informational quality. This approach directly addresses concerns regarding uncritical acceptance of AI outputs by positioning students as active evaluators rather than passive consumers.

Several institutional strategies have also emerged, including AI literacy curricula (Holmes & Tuomi, 2022), redesigned assessment practices that emphasize process documentation (Kasneci et al., 2023), and collaborative human-AI assignment designs (Yang et al., 2023). However, systematic evaluation of these approaches remains limited, with few studies employing robust experimental designs to assess their efficacy in preserving critical thinking development.

This literature review reveals a complex landscape of theoretical concerns, emerging empirical evidence, and potential pedagogical approaches. The current research builds upon these foundations by systematically synthesizing empirical findings to provide a more comprehensive understanding of how AI integration affects critical thinking development and which pedagogical strategies most effectively support cognitive skill development in AI-enriched educational environments.

3. Methodology

3.1 Research Design

This study employed a mixed-methods systematic review approach to synthesize findings from empirical research examining relationships between generative AI usage and critical thinking development. The mixed-methods design allowed integration of quantitative effect size calculations with qualitative thematic analysis of implementation contexts and pedagogical approaches. This approach aligns with methodological recommendations by Petticrew and Roberts (2006) for addressing complex educational interventions with heterogeneous outcome measures and implementation contexts.

3.2 Search Strategy and Selection Criteria

A comprehensive search strategy was developed to identify relevant empirical studies published between January 2010 and February 2025. The following electronic databases were systematically searched: ERIC, PsycINFO, Web of Science, Scopus, IEEE Xplore, ACM Digital Library, and Education Source. The search strategy employed combinations of terms related to AI technologies (e.g., "generative AI," "language models," "artificial intelligence," "machine learning"), educational applications (e.g., "education," "learning," "teaching," "pedagogy"), and critical thinking outcomes (e.g., "critical thinking," "analytical skills," "cognitive development," "reasoning abilities"). Additionally, citation tracking and reference list scanning of included studies were conducted to identify relevant publications potentially missed in database searches.

Studies were selected based on the following inclusion criteria:

1. Empirical studies with primary data collection (quantitative, qualitative, or mixed-methods)
2. Focus on generative AI applications in educational contexts
3. Measurement of critical thinking outcomes using validated instruments or systematic qualitative assessment
4. Publication in peer-reviewed academic journals ranked in Q1 or Q2 quartiles based on Scimago Journal Rankings
5. Publication between January 2010 and February 2025
6. English language publication

Studies were excluded if they:

1. Presented theoretical frameworks without empirical data
2. Focused exclusively on technical aspects of AI without educational applications
3. Measured only affective or engagement outcomes without addressing critical thinking
4. Examined non-generative AI applications exclusively
5. Provided insufficient methodological detail to assess quality

3.3 Data Extraction and Quality Assessment

Data extraction was conducted using a standardized form developed specifically for this review. Two researchers independently extracted data from each included study, with discrepancies resolved through discussion and consensus. Extracted information included:

1. Study characteristics (authors, publication year, journal, geographical location)
2. Methodological approach (design, sample size, duration, analytical methods)
3. Participant characteristics (educational level, age range, discipline)
4. AI intervention characteristics (type of AI system, implementation approach, frequency of use)
5. Critical thinking outcome measures (instruments used, dimensions assessed)
6. Key findings (effect sizes, qualitative themes, moderating factors)
7. Pedagogical approaches and implementation strategies

Methodological quality was assessed using the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018), which provides specific quality criteria for quantitative, qualitative, and mixed-methods studies. Each study received a quality rating ranging from 0 to 5, with higher scores indicating stronger methodological rigor. Studies were not excluded based on quality ratings; however, methodological limitations were considered in the synthesis and interpretation of findings.

3.4 Data Analysis

3.4.1 Quantitative Synthesis

For quantitative outcomes, effect sizes were calculated to standardize findings across studies. Hedges' g was used as the primary effect size measure, with positive values indicating beneficial effects of AI integration on critical thinking outcomes and negative values indicating detrimental effects. When studies reported multiple critical thinking outcomes, effect sizes were calculated separately for each dimension and then averaged to produce a composite effect size. Moderator analyses were conducted to examine how effect sizes varied based on:

1. Educational level (K-12, undergraduate, graduate)
2. Implementation approach (passive consumption vs. active engagement)
3. Duration of exposure (short-term, medium-term, long-term)
4. Specific critical thinking dimensions (analysis, evaluation, inference, etc.)

Random-effects meta-analysis was conducted using Comprehensive Meta-Analysis software (Version 3.0) to account for both within-study and between-study variance. Heterogeneity was assessed using the I^2 statistic, with values of 25%, 50%, and 75% interpreted as low, moderate, and high heterogeneity, respectively. Publication bias was evaluated using funnel plots and Egger's test.

3.4.2 Qualitative Synthesis

Qualitative data were analyzed using thematic synthesis (Thomas & Harden, 2008), which involved three stages: line-by-line coding of primary studies' findings, development of descriptive themes, and generation of analytical themes. This process facilitated identification of patterns across studies regarding contextual factors influencing AI's impact on critical thinking and effective pedagogical strategies. Coding was conducted independently by two researchers using NVivo 14 software, with regular meetings to resolve discrepancies and refine the coding framework.

3.4.3 Mixed-Methods Integration

Findings from quantitative and qualitative analyses were integrated using a convergent design framework (Creswell & Clark, 2017). This integration process involved:

1. Matching quantitative effect sizes with qualitative implementation contexts
2. Developing joint displays that visually represented relationships between implementation approaches and critical thinking outcomes
3. Identifying patterns of convergence and divergence between quantitative and qualitative findings
4. Developing integrated conclusions addressing the research questions

This mixed-methods approach allowed for a more comprehensive understanding of not only whether AI integration affects critical thinking development but also how and under what conditions these effects manifest.

4. Results

4.1 Characteristics of Included Studies

The systematic search process initially identified 1,473 potentially relevant citations. After removing duplicates and screening titles and abstracts, 214 full-text articles were assessed for eligibility. Following application of inclusion and exclusion criteria, 47 studies were included in the final review. Figure 1 presents the PRISMA flow diagram detailing the study selection process.

The 47 included studies comprised 28 quantitative studies, 12 qualitative studies, and 7 mixed-methods studies. Studies were conducted across 18 countries, with the largest number from the United States (n = 17), followed by China (n = 6), the United Kingdom (n = 5), Australia (n = 4), and other European and Asian countries (n = 15). Sample sizes ranged from 23 to 872 participants, with a total of 8,952 students across all studies. Table 1 presents characteristics of included studies.

Table 1. Characteristics of Included Studies (N = 47)

Characteristic	Number of Studies	Percentage
Study Design		
Experimental	19	40.4%
Quasi-experimental	9	19.1%
Longitudinal	11	23.4%
Case study	8	17.0%
Educational Level		
K-12	12	25.5%
Undergraduate	26	55.3%
Graduate	7	14.9%
Mixed levels	2	4.3%

Characteristic	Number of Studies	Percentage
Discipline		
Humanities	13	27.7%
Social Sciences	10	21.3%
STEM fields	17	36.2%
Interdisciplinary	7	14.9%
Study Duration		
< 1 month	13	27.7%
1-6 months	21	44.7%
> 6 months	13	27.7%
AI Implementation		
Instructor-mediated	22	46.8%
Student self-directed	14	29.8%
Hybrid approach	11	23.4%
MMAT Quality Score		
5 (Highest)	9	19.1%
4	17	36.2%
3	13	27.7%
2	6	12.8%
1 (Lowest)	2	4.3%

Critical thinking was assessed using various instruments across studies. The most commonly used standardized measures included the Watson-Glaser Critical Thinking Appraisal (n = 9), California Critical Thinking Skills Test (n = 7), and Cornell Critical Thinking Test (n = 5). Other studies employed course-specific assessments (n = 11), rubric-based evaluation of written work (n = 8), and qualitative assessment through interviews or think-aloud protocols (n = 7).

4.2 Quantitative Findings: Impact of AI Integration on Critical Thinking Dimensions

Meta-analysis of effect sizes from 35 studies providing sufficient quantitative data revealed a complex pattern of effects across different dimensions of critical thinking. The overall pooled effect size was g = -0.14 (95% CI [-0.27, -0.01]), indicating a small negative effect of generative AI integration on composite critical thinking outcomes. However, significant heterogeneity was observed ($I^2 = 76.3\%$, $p < .001$), suggesting substantial variation in effects across studies. Table 2 presents effect sizes for overall critical thinking and specific dimensions.

Table 2. Meta-Analysis Results for Critical Thinking Outcomes

Outcome	Number of Studies	Hedges' g	95% CI	p-value	I^2
Overall critical thinking	35	-0.14	[-0.27, -0.01]	.038	76.3%
Analysis skills	27	-0.31	[-0.46, -0.16]	<.001	68.7%
Evaluation skills	24	-0.22	[-0.38, -0.06]	.007	72.1%
Inference skills	21	0.05	[-0.13, 0.23]	.581	74.6%
Problem-solving	18	0.19	[0.04, 0.34]	.013	65.9%
Argumentation	16	-0.28	[-0.45, -0.11]	.001	70.3%

The analysis revealed varying impacts across different critical thinking dimensions. The largest negative effects were observed for analysis skills (g = -0.31) and argumentation (g = -0.28), while problem-solving showed a small positive effect (g = 0.19), and inference skills demonstrated no significant effect (g = 0.05). This pattern suggests that AI integration may affect different cognitive processes in distinct ways rather than uniformly enhancing or diminishing critical thinking capacities.

Moderator analyses revealed several significant factors influencing the relationship between AI integration and critical thinking outcomes. Table 3 presents results of moderator analyses examining how effect sizes varied across different implementation contexts.

Table 3. Moderator Analysis Results

Moderator Variable	Category	N	Hedges' g	95% CI	p-value
Educational Level	K-12	9	-0.08	[-0.29, 0.13]	.459
	Undergraduate	20	-0.16	[-0.32, 0.00]	.049
	Graduate	6	-0.17	[-0.41, 0.07]	.168
Implementation Approach	Passive consumption	14	-0.37	[-0.53, -0.21]	<.001
	Active engagement	21	0.08	[-0.05, 0.21]	.226

Moderator Variable	Category	N	Hedges' g	95% CI	p-value
Duration	< 1 month	11	-0.06	[-0.25, 0.13]	.538
	1-6 months	15	-0.14	[-0.31, 0.03]	.107
	> 6 months	9	-0.24	[-0.45, -0.03]	.025
AI System Usage	Content generation	17	-0.31	[-0.47, -0.15]	<.001
	Interactive dialogue	10	0.12	[-0.07, 0.31]	.227
	Combined approaches	8	-0.09	[-0.30, 0.12]	.392

The most significant moderator was implementation approach ($Qb = 19.47$, $p < .001$), with passive consumption of AI-generated content associated with moderate negative effects on critical thinking ($g = -0.37$), while active engagement approaches showed a small, non-significant positive effect ($g = 0.08$). Duration of AI exposure also emerged as a significant moderator ($Qb = 6.83$, $p = .033$), with longer exposure periods associated with larger negative effects.

The specific usage pattern of AI systems significantly moderated outcomes ($Qb = 13.26$, $p = .001$), with content generation applications showing negative effects ($g = -0.31$) and interactive dialogue applications showing small, non-significant positive effects ($g = 0.12$). Educational level did not emerge as a significant moderator ($Qb = 3.21$, $p = .200$), suggesting similar patterns across K-12, undergraduate, and graduate contexts, though effects reached statistical significance only at the undergraduate level.

4.3 Qualitative Findings: Implementation Contexts and Mechanisms

Thematic synthesis of qualitative findings revealed four major themes related to how AI implementation contexts shape critical thinking outcomes: (1) passive versus active engagement modalities, (2) metacognitive scaffolding, (3) feedback dynamics, and (4) assessment alignment. Table 4 summarizes these themes with representative findings from primary studies.

Table 4. Qualitative Themes Regarding Implementation Contexts

Theme	Description	Representative Findings
Passive vs. Active Engagement	Distinction between using AI as content provider versus analytical partner	"Students who used AI primarily to generate finished products showed diminished engagement with analytical processes, while those using AI as a dialogue partner demonstrated enhanced reflection." (Thompson et al., 2023)
Metacognitive Scaffolding	Presence of explicit guidance for evaluating and reflecting on AI interactions	"Courses incorporating explicit instruction in evaluating AI outputs and identifying reasoning flaws showed preserved critical thinking development despite high AI usage." (Diakopoulos, 2023)

Theme	Description	Representative Findings
Feedback Dynamics	Nature of feedback processes between students, instructors, and AI systems	"When AI feedback replaced rather than supplemented instructor feedback, students showed decreased ability to recognize conceptual errors and logical inconsistencies." (Martinez-Maldonado et al., 2022)
Assessment Alignment	Congruence between assessment methods and AI integration approach	"Assessments emphasizing process documentation and evolving analytical thinking showed stronger alignment with constructive AI integration compared to outcome-focused evaluations." (Wang & Koedinger, 2023)

The qualitative analysis revealed several mechanisms through which AI integration appeared to influence critical thinking development. In contexts characterized by passive consumption, students reported decreased motivation to engage in independent analytical processes, consistent with cognitive offloading theories. As one participant in Yeung et al.'s (2023) study noted: "When I know I can get the AI to write an analysis, I spend less time thinking deeply about the material myself."

Conversely, active engagement approaches emphasized AI as a dialogic partner rather than a content provider. In these contexts, students reported increased metacognitive awareness, as evidenced by this reflection from a participant in Thompson et al.'s (2023) study: "Comparing my reasoning to the AI's approach helped me identify gaps in my own thinking process that I wouldn't have recognized otherwise."

The presence of explicit metacognitive scaffolding emerged as a crucial mediating factor. Studies incorporating structured protocols for evaluating AI outputs and reflecting on reasoning processes showed preserved or enhanced critical thinking development despite high AI usage levels. These approaches directly addressed potential overreliance by positioning students as critical evaluators rather than passive consumers of AI-generated content.

4.4 Integration of Quantitative and Qualitative Findings

The integration of quantitative and qualitative findings revealed several patterns of convergence. First, both analyses identified implementation approach as the most significant factor influencing critical thinking outcomes, with passive consumption associated with negative effects and active engagement showing neutral or positive effects. Second, both strands highlighted differential impacts across critical thinking dimensions, with analysis and argumentation skills showing greater vulnerability to attenuation compared to problem-solving and inference skills.

The mixed-methods integration also revealed that negative effects were most pronounced in contexts lacking explicit metacognitive scaffolding, regardless of the specific AI system employed. This finding suggests that pedagogical approach may be more influential than technological characteristics in determining critical thinking outcomes.

Figure 2 presents a conceptual model synthesizing these integrated findings, illustrating how implementation approach and pedagogical factors mediate the relationship between AI integration and critical thinking outcomes. The model identifies specific pathways through which different engagement modalities influence distinct critical thinking dimensions.

4.5 Pedagogical Strategies with Empirical Support

Analysis of implementation approaches across studies with neutral or positive outcomes identified several pedagogical strategies with empirical support for preserving critical thinking development in AI-integrated environments. Table 5 summarizes these strategies with supporting evidence from included studies.

Table 5. Evidence-Based Pedagogical Strategies for AI Integration

Strategy	Description	Supporting Evidence
AI-Augmented Critical Dialogues	Structured analytical conversations between students and AI systems with explicit reflection	Martinez-Maldonado et al. (2022): Implementation with 147 undergraduate students showed significant improvements in analytical writing compared to control condition (g = 0.38)
Comparative Analysis Exercises	Tasks requiring students to compare human-generated and AI-generated analyses of the same material	Ramirez et al. (2024): Students engaging in comparative analysis showed 27% higher gains in independent critical analysis compared to direct consumption group
Process Documentation Requirements	Assignments requiring documentation of analytical steps rather than only finished products	Wang & Koedinger (2023): Process documentation requirements mitigated negative effects of AI usage on analysis skills (g = -0.07 vs. g = -0.34)
AI Literacy Curriculum	Explicit instruction in AI capabilities, limitations, and evaluation strategies	Diakopoulos (2023): Integration of AI literacy modules associated with preserved critical thinking gains despite high AI usage (g = 0.12)
Error Analysis Tasks	Activities requiring students to identify and correct reasoning flaws in AI-generated content	Thompson et al. (2023): Error analysis activities associated with 23% improvement in evaluation skills compared to baseline

These strategies share several common elements: they position students as active evaluators rather than passive consumers, incorporate explicit metacognitive reflection, and emphasize process over product. The strongest empirical support was found for approaches combining multiple strategies rather than implementing single interventions in isolation.

5. Discussion

5.1 Interpretation of Main Findings

This systematic review reveals a complex relationship between generative AI integration and critical thinking development that defies simplistic characterization as uniformly beneficial or detrimental. The small negative overall effect size (g = -0.14) must be interpreted within the context of significant heterogeneity and substantial moderating factors, particularly implementation approach and specific critical thinking dimensions.

The findings suggest that potential erosion of critical thinking skills is not an inherent consequence of AI integration but rather emerges primarily in contexts characterized by passive consumption modalities, particularly when students use AI systems principally as content generators without engaging in analytical processes themselves. This pattern aligns with cognitive offloading theories suggesting that externalizing cognitive processes to technological tools may attenuate skill development through reduced practice (Risko & Gilbert, 2016).

The differential impact across critical thinking dimensions—with analysis and argumentation showing vulnerability to attenuation while problem-solving demonstrated enhancement—suggests that AI integration may reshape rather than uniformly diminish cognitive skill development. This pattern may reflect the specific capabilities of contemporary generative AI systems, which excel at producing structured analyses and coherent arguments but require human guidance for novel problem formulation (Yang et al., 2023).

Perhaps most significantly, the findings demonstrate that implementation approach substantially mediates outcomes, with active engagement modalities showing neutral or positive effects despite high levels of AI usage. This suggests that the critical variable is not whether students use AI tools but how they engage with these systems—as content providers versus analytical partners. The strong moderating effect of implementation approach (passive consumption g = -0.37 vs. active engagement g = 0.08) represents the largest effect size difference identified in the analysis, underscoring its central importance.

The duration effect—with longer exposure periods associated with larger negative effects—raises important questions about adaptation processes and potential cumulative impacts. This finding may reflect gradual shifts in cognitive habits and analytical approaches through sustained engagement with AI systems, consistent with theoretical perspectives on technological mediation (Verbeek, 2015). However, this interpretation must be tempered by the observation that longer-duration studies often involved less structured implementation approaches, potentially confounding these variables.

5.2 Implications for Educational Practice

The findings have several important implications for educational practice in increasingly AI-integrated learning environments. First, they suggest that educators should focus less on whether to incorporate AI tools and more on how to structure their implementation to preserve critical thinking development. The evidence supports approaches that position students as active evaluators rather than passive consumers of AI-generated content.

Second, the research identifies specific pedagogical strategies with empirical support for mitigating potential negative effects, including AI-augmented critical dialogues, comparative analysis exercises, process documentation requirements, and error analysis tasks. These approaches directly address potential overreliance by encouraging metacognitive reflection and positioning AI as a complement to rather than substitute for human analytical processes.

Third, the findings suggest that assessment practices require recalibration to align with AI-integrated learning environments. Traditional assessment approaches emphasizing polished final products may inadvertently encourage passive consumption of AI-generated content. In contrast, process-oriented assessments that document analytical development, require comparative evaluation, or incorporate metacognitive reflection appear more congruent with constructive AI integration (Wang & Koedinger, 2023).

Fourth, the differential vulnerability of specific critical thinking dimensions suggests that targeted interventions may be necessary to preserve analytical and argumentative capabilities. For instance, explicit instruction and practice in constructing original arguments may become increasingly important as generative AI systems excel at producing well-structured argumentation that students might otherwise passively consume (Yang et al., 2023).

Finally, the findings highlight the importance of developing comprehensive AI literacy curricula that extend beyond technical understanding to include critical evaluation of AI outputs, recognition of reasoning limitations, and metacognitive awareness of human-AI collaborative processes. Such curricula may represent an essential educational response to increasingly sophisticated AI systems (Diakopoulos, 2023; Holmes & Tuomi, 2022).

5.3 Theoretical Implications

The findings contribute to theoretical understanding of how technological mediation influences cognitive development in several ways. First, they provide empirical support for cognitive offloading theories by demonstrating potential attenuation of specific analytical skills through externalization of cognitive processes. However, they also suggest important boundary conditions for these effects, with active engagement modalities showing preserved skill development despite high AI usage (Risko & Gilbert, 2016).

Second, the findings align with technological mediation theory by demonstrating that AI tools do not simply augment existing cognitive processes but potentially transform how students engage with information and develop analytical capabilities (Verbeek, 2015). The differential effects across critical thinking dimensions suggest that rather than uniform enhancement or diminishment, AI integration may reshape the developmental trajectory of cognitive skills in nuanced ways.

Third, the findings extend social constructivist perspectives by illuminating how AI systems function within collaborative meaning-making processes. The evidence suggests that when positioned as dialogic partners rather than authoritative knowledge sources, AI systems can contribute productively to cognitive development through scaffolded analytical interactions (Vygotsky, 1978; Holmes & Tuomi, 2022).

5.4 Limitations and Future Research Directions

This review has several limitations that should inform interpretation of the findings and guide future research. First, despite efforts to identify all relevant studies, the rapidly evolving nature of AI technologies means that some recent implementations may be underrepresented in the published literature. Second, the heterogeneity in how critical thinking was conceptualized and measured across studies limits direct comparability, though standardized effect size calculations mitigate this concern. Third, many studies relied on self-reported measures or assessments conducted in controlled environments, potentially limiting ecological validity for understanding impacts in naturalistic settings.

Several promising directions for future research emerge from these findings. First, longitudinal studies with standardized critical thinking assessments are needed to better understand developmental trajectories and potential adaptation processes over extended periods. These studies should incorporate multiple measurement time points to capture nonlinear effects and track different critical thinking dimensions separately.

Second, experimental studies systematically comparing different implementation approaches within the same educational contexts would provide stronger causal evidence regarding effective pedagogical strategies. These studies should control for technological characteristics while manipulating implementation variables such as engagement modality, metacognitive scaffolding, and feedback processes.

Third, qualitative investigations of students' experiences with different AI engagement modalities would enhance understanding of cognitive processes that may not be captured by standardized assessments. These studies could employ think-aloud protocols, reflective journals, or ethnographic methods to illuminate how students navigate analytical challenges in AI-integrated environments.

Fourth, cross-disciplinary research examining whether effects vary across knowledge domains would address an important gap in current understanding. Such research should explore whether discipline-specific analytical practices interact with AI integration in ways that influence critical thinking development distinctively across fields.

Finally, comprehensive evaluation of institutional responses such as AI literacy curricula, assessment redesign initiatives, and faculty development programs would provide valuable guidance for educational policy and practice. These evaluations should examine both implementation processes and outcomes to identify factors influencing effectiveness across diverse educational contexts.

6. Conclusion

This systematic review synthesized empirical evidence regarding the relationship between generative AI integration in educational contexts and critical thinking development. The findings reveal a complex relationship mediated by implementation approach, specific critical thinking dimensions, and pedagogical factors. While passive consumption of AI-generated content shows moderate negative associations with analytical and argumentative skills, active engagement approaches demonstrate neutral or positive effects even with high levels of AI usage.

These findings challenge simplistic narratives regarding AI's impact on cognitive development, suggesting that potential erosion of critical thinking skills is not an inevitable consequence of technological integration but rather emerges primarily through specific modalities of engagement. The substantial moderating effect of implementation approach—with passive consumption and active engagement showing divergent outcomes—underscores the central importance of pedagogical decisions in shaping how AI technologies influence cognitive development.

The research identifies several evidence-based pedagogical strategies for preserving critical thinking development in AI-integrated environments, including AI-augmented critical dialogues, comparative analysis exercises, process documentation requirements, AI literacy instruction, and error analysis tasks. These approaches share common elements of positioning students as active evaluators rather than passive consumers and incorporating explicit metacognitive reflection on analytical processes.

As generative AI technologies continue to evolve and proliferate throughout educational contexts, this research suggests that educators should focus not on whether these tools should be integrated but rather on how their implementation can be structured to preserve and enhance critical cognitive capabilities. By approaching AI as a potential partner in analytical processes rather than a substitute for human cognition, educational practices can leverage technological affordances while preserving the developmental trajectories essential for critical thinking.

Future research should address remaining gaps through longitudinal studies tracking developmental trajectories, experimental investigations of specific pedagogical approaches, qualitative explorations of student experiences, cross-disciplinary comparisons, and comprehensive evaluations of institutional responses. These efforts will further refine understanding of how educational practices can adapt to technological change while preserving essential cognitive capabilities in an increasingly AI-integrated world.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. Review of Educational Research, 85(2), 275-314. https://doi.org/10.3102/0034654314551063

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of Artificial Intelligence and Education, 3(1), 14-36. https://doi.org/10.38126/JAIDE030102

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623). https://doi.org/10.1145/3442188.3445922

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... & Amodei, D. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems, 33, 1877-1901.

Chen, X., Li, J., Zhang, R., & Wang, Y. (2024). Artificial intelligence in higher education: A longitudinal study of student usage patterns and academic outcomes. Computers & Education, 201, 104978. https://doi.org/10.1016/j.compedu.2023.104978

Creswell, J. W., & Clark, V. L. P. (2017). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Davies, M., & Barnett, R. (Eds.). (2015). The Palgrave handbook of critical thinking in higher education. Palgrave Macmillan. https://doi.org/10.1057/9781137378057

Diakopoulos, N. (2023). Thinking critically about AI: Developing student competencies for evaluating artificial intelligence in education. Journal of Media Literacy Education, 15(2), 45-62. https://doi.org/10.23860/JMLE-2023-15-2-4

Facione, P. A. (2011). Critical thinking: What it is and why it counts. Insight Assessment, 1(1), 1-28.

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(4), 542-570. https://doi.org/10.1111/ejed.12533

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., & Vedel, I. (2018). Mixed methods appraisal tool (MMAT), version 2018. Registration of Copyright (#1148552), Canadian Intellectual Property Office.

Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. Review of Educational Research, 86(2), 431-468. https://doi.org/10.3102/0034654315605917

Ihde, D. (1990). Technology and the lifeworld: From garden to earth. Indiana University Press.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. ETS Research Report Series, 2014(1), 1-23. https://doi.org/10.1002/ets2.12009

Martinez-Maldonado, R., Buckingham Shum, S., Pardo, A., Hernandez-Leo, D., Kitto, K., Bartindale, T., Chen, G., Echeverria, V., Elliott, D., Lobzhanidze, A., Martinez-Ormiston, M., Prieto-Alvarez, C., & Tsai, Y. (2022). AI-augmented critical pedagogy: Negotiating human-AI partnerships in educational dialogues. British Journal of Educational Technology, 53(6), 1713-1733. https://doi.org/10.1111/bjet.13273

Marzano, R. J., & Heflebower, T. (2012). Teaching & assessing 21st century skills. Solution Tree Press.

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Blackwell Publishing. https://doi.org/10.1002/9780470754887

Ramirez, C., Hinojosa, C., & Rodriguez, M. (2024). Effects of AI-assisted writing on analytical skill development: An experimental study with high school students. Journal of Writing Research, 15(3), 312-339. https://doi.org/10.17239/jowr-2024.15.03.04

Rear, D. (2019). One size fits all? The limitations of standardised assessment in critical thinking. Assessment & Evaluation in Higher Education, 44(5), 664-675. https://doi.org/10.1080/02602938.2018.1526255

Reich, J., Kim, Y., Robinson, C., Roy, D., & Thompson, M. (2023). AI, education, and the automation paradox. Science, 380(6647), 825-827. https://doi.org/10.1126/science.adi9963

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8, 45. https://doi.org/10.1186/1471-2288-8-45

Thompson, K., Martinez-Maldonado, R., & Buckingham Shum, S. (2023). AI as partner or substitute? Contrasting approaches to generative AI integration and their impact on analytical writing development. International Journal of Artificial Intelligence in Education, 33(4), 752-783. https://doi.org/10.1007/s40593-023-00338-1

Verbeek, P. P. (2015). Beyond interaction: A short introduction to mediation theory. Interactions, 22(3), 26-31. https://doi.org/10.1145/2751314

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wang, L., & Koedinger, K. R. (2023). From cognitive offloading to cognitive partnership: A three-year longitudinal study of AI integration in undergraduate education. npj Science of Learning, 8(1), 7. https://doi.org/10.1038/s41539-023-00156-z

Yang, Z., Wei, Z., Ye, Q., Wu, J., Zhou, H., Zhou, Z., He, X., Nguyen, T., Akyürek, A. F., Nachum, O., Tenenbaum, J. B., Gan, C., Zhou, B., & Liang, P. (2023). Foundation models for decision making: Problems, methods, and opportunities. ArXiv Preprint. https://doi.org/10.48550/arXiv.2303.04129

Yeung, A. W. K., Tosevska, A., Klager, E., Eibensteiner, F., Laxar, D., Stoyanov, J., Glisic, M., Zeiner, S., Kulnik, S. T., Crutzen, R., Kimberger, O., Kletecka-Pulker, M., Atanasov, A. G., & Willschke, H. (2023). Medical and biomedical students' perspectives of ChatGPT and other large language models. Journal of Medical Internet Research, 25(4), e45826. https://doi.org/10.2196/45826

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J. B., Yuan, J., & Li, Y. (2022). A review of artificial intelligence (AI) in education from 2010 to 2020. Complexity, 2022, 1-18. https://doi.org/10.1155/2022/8143334

Zhang, T., Wu, Y. J., & Cui, B. (2023). Exploring the impact of ChatGPT on education: Applications, challenges, and future research directions. Educational Technology & Society, 26(3), 56-69.

Zhou, J., Park, C. Y., Maddock, J., Li, R., Zhou, C., Haggins, A., Chu, L., Golley, P., Redmond, P., Santacroce-Tejedor, R., Qi, G., He, X., & Qureshi, A. (2023). Large language models for medical education: Performance of ChatGPT across multiple choice medical exam questions. PLOS Digital Health, 2(6), e0000225. https://doi.org/10.1371/journal.pdig.0000225

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books.

………………………………..