FACULTAD DE CIENCIAS HUMANAS Y

SOCIALES

Assessment of AI's Impact on Critical Thinking in Higher Education: A Systematic Review

Author: Pedrosa Prats, Elena

Supervisor: Vendrell Morancho, Mireia

Madrid

2024/2025

Abstract

Background: The integration of artificial intelligence (AI) tools in higher education has accelerated rapidly, prompting growing interest in their impact on students' critical thinking (CT) skills. While some studies suggest that AI can enhance analytical reasoning and reflective learning, others raise concerns about overreliance and cognitive offloading. Despite this growing literature, the conceptualization and measurement of "critical thinking" remain inconsistent across studies, complicating efforts to synthesize findings. **Objectives:** This systematic review examines how recent empirical studies (2022–2025) define, operationalize, and assess critical thinking in the context of AI-enhanced learning in higher education. Specifically, it investigates the theoretical frameworks employed, the assessment tools used, the types of AI tools integrated, and the reported outcomes on CT development. **Methods:** Following the PRISMA 2020 guidelines, we conducted a systematic search in Scopus, Web of Science, and ERIC for studies published between 2022 and February 2025 that involved higher education student populations, AI-based tools, and CT-related outcomes. A total of 22 eligible studies were identified and analyzed using narrative synthesis and thematic coding. Risk of bias was assessed using JBI and CASP tools**. Results:** Only 8 of the 22 studies provided a formal definition of critical thinking, and even fewer used dedicated CT assessment instruments. Most studies relied on mixed methods and domain-specific performance tasks. The findings indicate that AI tools can support CT development, particularly when embedded in human-facilitated learning environments that promote reflection, evaluation, and dialogue. However, studies also reported risks such as superficial learning and diminished metacognitive engagement when AI was used as a cognitive substitute. **Conclusions:** AI's impact on critical thinking in higher education is shaped by tool design, instructional context, and the clarity of CT conceptualization. This review highlights the need for consistent definitions, theoretically grounded assessments, and pedagogical models that combine AI affordances with reflective, instructor-guided learning. Future research should emphasize longitudinal designs and the development of CT-specific instruments aligned with validated frameworks.

Keywords: Critical thinking, artificial intelligence, higher education, ChatGPT, systematic review, AI in education, PRISMA 2020, cognitive skills, metacognition

Table of Contents

Introduction

Definitions and Theoretical Frameworks of Critical Thinking

John Dewey is frequently credited as an early theorist of critical thinking, referring to it as "reflective thinking”, which stands for the active, persistent, and careful consideration of a belief or knowledge claim in light of the grounds that support it. Since then, many influential frameworks have emerged, all agreeing that critical thinking skills involve higher-order thinking (Smith, 2020; Aktoprak & Hürsen, 2022).

One of the most widely cited is the **Delphi consensus** project led by (Facione, 1990), which states CT must encompass dispositions and skills, one alone does not ensure critical thinking in practice. Seen as a form of purposeful, self-regulatory judgment involving dispositions (such as willingness to question assumptions, open-mindedness, intellectual curiosity, fair-mindedness, or skepticism of unsubstantiated claims) that reflect a habitual mindset that encourages the application of the following core cognitive skills: interpretation, analysis, evaluation, inference, explanation, and self-regulation.

Halpern (1998) further emphasized that critical thinking involves the **use of cognitive strategies** aimed at increasing the likelihood of desirable outcomes. Her framework underscores goal-directed, reflective thought processes and the application of logic to real-world problems. Similarly, Ennis (1987, 2011) defined CT as "reasonable, reflective thinking focused on deciding what to believe or do,” integrating both cognitive rigor and a disposition toward evidence-based reasoning.

Other notable contributions include the **Paul and Elder (2012)** model, which defines CT as the active process of analyzing and evaluating thinking with the goal of improving it. According to their view, CT is metacognition with the explicit intention of making thinking more disciplined and effective.

Across these frameworks, there is a shared emphasis on critical thinking as a **higher-order cognitive activity** that involves both **reasoning skills** and **dispositional habits of mind**. Yet subtle differences exist in scope and emphasis. For instance, whether critical thinking encompasses creativity or moral reasoning, and whether it is best taught as a general skill or within domain-specific contexts (Moseley et al., 2005). Most universities have adopted a blended approach, embedding CT instruction within discipline-specific courses rather than teaching it in isolation.

Study Rationale and Aims

Over the past two decades, educational research on critical thinking (CT) has grown steadily, with a marked surge in 2023 reflecting the increasing focus on the intersection between emerging technologies and CT (Yücel, 2025; Walter, 2024; Melisa et al., 2025). This surge coincides with the rapid integration of artificial intelligence (AI) tools into higher education, where technology-driven approaches to learning are receiving growing academic attention and citation (Yücel, 2025; Walter, 2024; Melisa et al., 2025). New AI applications continue to emerge at an accelerated pace, each varying in functionality, capabilities, and educational effects (Rai, 2024; Sasikala & Ravichandran, 2024; Melisa et al., 2025).

The convergence of AI and CT in higher education is a recent development, catalyzed by the public release of generative AI tools such as ChatGPT in late 2022 (Rahyuni et al., 2025). Despite the growing body of literature, no systematic review has yet focused specifically on the **conceptual foundations** of CT within AI-mediated learning. This omission is particularly significant given the definitional ambiguity that surrounds both constructs and the diversity of approaches used to assess CT.

As interest intensifies, so too do contrasting perspectives: while some argue that AI-powered assistants may scaffold and deepen students' reasoning processes, others raise concerns that such tools may foster passivity and superficial engagement. Moreover, without a shared definition of CT, it becomes difficult to meaningfully compare outcomes across studies.

This systematic review addresses these gaps by analyzing recent empirical research on AI's impact on critical thinking in higher education. Its primary goal is to clarify how CT is being **conceptualized and operationalized** in these studies, and what outcomes are being reported. Specifically, the review asks:

- How have recent studies conceptualized or defined CT when integrating AI in higher education?
- What is being measured and understood as critical thinking in studies on AI’s impact on higher education students’’ critical thinking?

- What types of AI tools and interventions are being implemented, and for what educational purposes?
- What impacts on students' CT skills (positive, neutral, or negative) are reported as a result of using AI?

In summary, this review responds to an urgent need for conceptual clarity at the intersection of AI and critical thinking. By mapping the definitions and measurements used across recent studies, it seeks to bring coherence to a rapidly evolving field. Clarifying how CT is framed in AI-integrated education will not only aid interpretation of current findings but also inform future research designs and pedagogical practices. Ultimately, when we discuss "critical thinking in the age of AI," we must ensure that we are speaking a **shared and well-defined language**. This review sets the foundation for that conversation.

Methods

3.1 Review Protocol and Registration

For this systematic review to be conducted, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines were followed to provide a standard peer-accepted methodology that contributes to the quality assurance of the revision process and ensures its replicability (Page et al., 2021). The protocol was developed prior to the review process and adhered strictly to the predefined eligibility criteria and search methods.

The review was not registered in PROSPERO or another international review registry, but all procedures and criteria are fully reported to ensure transparency and replicability.

3.2 Eligibility Criteria

The eligibility criteria were defined using the PICOS framework (Population, Intervention, Comparator, Outcomes and Study design) and are summarized in Table 1. These criteria were designed to ensure inclusion of peer.reviewed, empirical research focused on the impact of AI on CT within higher education contexts.

Studies published between January 2022 and February 2025 were included to reflect the most recent phase of educational research on AI and critical thinking. This period was selected because 2022 marked the public release of widely used generative AI tools (e.g., ChatGPT), and 2023 saw their rapid adoption in environments in higher education, accompanied by a surge in research exploring their impact on critical thinking (Yücel, 2025; Mustafa et al., 2024; Sasikala & Ravichandran, 2024).

Moreover, to balance the scope of this review while ensuring the result's reliability with an available large pool of data to draw valid conclusions, the population was intentionally selected to be students, specifically higher education students, which represent the majority of the body of literature on the topic (Mustafa et al., 2024).

3.2.1 & 3.2.2 Inclusion and Exclusion Criteria

Figure 1:
Inclusion and Exclusion Criteria for Systematic Review (PRISMA 2020)

Criterion	Included	Excluded
(1) Database	(1.1) Scopus, ERIC, and Web of Science	(1.2) Other databases
(2) Publication Year	(2.1) 2022–2025	(2.1) Before 2022
(3) Language	(3.1) English	(3.2) Other languages
(4) Publication Type & Peer-Review Status	(4.1) Peer-reviewed empirical journal articles	(4.2) Non-empirical work (e.g., editorials, opinion pieces, blogs, book chapters, editorials, theoretical discussions, etc.)
(5) Accessibility	(5.1) Full text available via open or institutional access	(5.2) Irretrievable after exhaustive efforts (e.g., contacting authors, libraries).
(6) Population	(6.1) Higher education students (e.g. undergraduate, graduate, etc.)	(6.2) Non-higher education students (e.g., K–12, corporate training, teachers, etc.).
(7) Study Focus	(7.1) Assesses AI's impact on CT (7.2) CT is a primary or secondary outcome (7.3) Describes the assessment method used to measure AI's impact on CT, whether through structured methods (quantitative, qualitative, or mixed-methods) or unstructured methods	(7.4) Mentions AI, CT and/or AI's role in education anecdotally (7.5) Does not specify the method of assessing AI's impact on CT (7.6) CT is not a measured outcome
(8) AI Intervention	(8.1) Clearly defines specific AI tools or technologies	(8.2) Does not mention specific AI tools or technologies

3.3 Information Sources

To ensure comprehensive coverage, three major academic database dwere selected:

- Scopus
- ERIC (Education Resources Information Center)
- Web of Science

These databases were chosen for their high relevance to education, learning technologies, and social sciences, and their inclusion of high-quality, peer-reviewed research on AI and CT. No additional grey literature sources were used to maintain a strict focus on peer-reviewed empirical studies.

3.4 Search Strategy

The final search was conducted on February 23rd, 2025. A targeted search strategy was developed using Boolean operators and keyword clusters tailored to each database. To maintain consistency across platforms, filters and fields were harmonized as much as possible, while acknowledging database-specific formatting.

The search string was designed to capture studies at the intersection of AI, CT, and higher education assessment, and was structured as follows:

Search string format:

TITLE-ABS-KEY ((*CT substring*) AND (*Education substring*) AND (*AI substring*)).

Substrings used:

CT substring: "Critical Thinking"

Education substring: "Higher Education" OR "University" OR "College" OR "Undergraduate*"

Methodology substring: "Assess*" OR "Evaluat*" OR "Measure*" OR "Rubric*" OR "Framework*" OR "Instrument*" OR "Tool*" OR "Test*"

AI substring: "Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning" OR "Generative AI" OR "Large Language Models" OR "LLM" OR "Natural Language Processing" OR "NLP" OR "Intelligent Tutoring Systems" OR "ITS" OR "AI Chatbot*" OR "Virtual Assistant*" OR "AI-Based Feedback" OR "AI-Powered Feedback" OR "AI-Assisted Feedback" OR "AI-Assisted Learning" OR "Automated Feedback" OR "Adaptive Learning AI" OR "Gamifi*"

Database-specific filters were applied for:

- Language (English)
- Publication Type (peer-reviewed journal articles)
- Publication Year (2022-2025)

All the specific filters used for each database are documented in Appendix E: PRISMA 2020 Checklist.

3.5 Selection Process

After completing the database search, all records were imported into Zotero for deduplication and screening. A single reviewer conducted the initial screening of titles and abstracts, using the predefined eligibility criteria (detailed in Table 1). Studies that clearly failed to meet inclusion criteria were excluded at this stage.

The remaining articles were retrieved in full-text and underwent independent screening by two reviewers. Discrepancies in inclusion decisions were resolved through discussion and consensus. This two-step selection process followed the PRISMA 2020 recommendations to ensure transparency and reproducibility.

3.6 Data Extraction

This process was performed independently by two reviewers using a pre-piloted standardized form developed in Microsoft Excel. The form was tested on a small sample of studies to ensure clarity and consistency. One of the reviewers extracted data from the included studies, and any discrepancies were resolved through mutual discussion with their colleague.

All extracted data were cross-checked for accuracy. Where essential information was missing or unclear, study authors were contacted via email to request clarification.

3.7 Data Items Extracted

The following data items were extracted for each included study, based on their relevance to the review's objectives:

- Study identification: Authors, year, title, and country.
- Study design and methodology: Type of study , darta type, sample size, and paritciapnt characteristics.
- AI tools studied: Specific applications of AI
- CT as an outcome: Whether CT was a primary or secondary outcome
- Aspects of CT measured: Cognitive dimensions targeted
- Definitions of CT: Whether the study explicitly defined CT, and how
- Key findings related to CT: Improvements, neutral effects, or declines in the aspects related to CT
- Qualitative insights (Where applicable): Themes or quotes relevant to student's perceptions, learning experiences or AI influence on CT

All data items were aligned with the research questions and review objectives and were used to uniform both the narrative synthesis and thematic analysis.

3.8 Risk of Bias Assessment

The risk of bias (RoB) in the included studies was assessed systematically to ensure transparency and to inform the interpretation of findings. The selection of tools and procedures was based on the study designs and the mixed methodological nature of the included literature.

A tailored approach was adopted to accommodate the range of empirical designs across the included studies (qualitative, quantitative and mixed methods). Two reviewers independently conducted the risk of bias assessment for all included studies, and discrepancies were resolved through discussion and consensus.

Quantitative and Mixed-Methods Studies

For studies with quantitative or quasi-experimental components, the Joanna Briggs Institute (JBI) Critical Appraisal Checklists were used. Depending on study design, the appropriate tool was selected:

- JBI Checklist for quasi-experimental studies
- JBI Checklist for analytical cross sectional studies
- JBI Checklist for randomized controlled trials (where applicable)

Each study was evaluated against the relevant checklist's items, and an overall RoB judgment was made (low, moderate, or high). The criteria evaluated included:

- Clarity of cause-effects relationship
- Confounding factors and how they were managed
- Consistency of outcome measures
- Validity and reliability of instruments
- Adequacy of follow-up and outcome reporting

Qualitative studies

For qualitative studies, the Critical Appraisal Skills Programme (CASP) checklist was used. This tool evaluates:

- Clarity of research aims and methodology
- Appropriateness of the research design
- Recruitment of sampling strategy
- Ethical considerations
- Rigor of analysis
- Transparency and coherence of findings

Each study was rated as having low, moderate or high risk of bias, baked on how many of the CASP criteria were met or partially met.

Synthesis use

The RoB assessments were not used as exclusion criteria but were instead used to:

- Interpret the confidence in individual study findings
- Inform the weighting of evidence in the narrative synthesis
- Highlight limitations and methodological concerns across the body of literature

The complete risk of bias evaluations are presented in Appendix D: Risk of Bias Tables, which include item-level assessments and summary judgments for each included study.

3.9 Data Synthesis Approach

Given the heterogeneity of study designs, AI tools, and CT outcome measured, a meta-analysis was not conducted. Instead, a narrative synthesis was used to systematically organize and interpret the findings across included studies.

The synthesis followed the general principles outlined by the ESRC Guidance on Narrative Synthesis and complied with PRISMA 2020 guidance for non-quantitative reviews. The results were structured thematically to identify patterns and divergences across studies. Particular attention was paid to:

- The types of AI tools and systems employed
- The assessment methods used to measure CT
- The specific dimensions of CT targeted
- The role of AI in either facilitating, supporting or assessing CT skills
- Reported outcomes: improvement, neutral effects, or challenges

The synthesis approach emphasized conceptual clarity, contextual relevance and methodological transparency.

3.9.1 Narrative Synthesis Methods

The narrative synthesis was conducted in four stages:

1. Preliminary mapping

   All included studies were charted using a standardized summary table to identify key study features (study design, AI tool, CT outcome type, instruments, etc.)

2. Developing a thematic framework

   Initial reading of all extracted data (qualitative and quantitative findings) led to the development of an a priori thematic structure based on the review questions. Thee themes included:

        - Definitions and conceptualizations of CT
        - Instruments and tools for assessing CT
        - CT dimensions targeted
        - Types of AI used (eg. chatbots, generative tools, intelligent tutoring systems)
        - Role of AI (assessment, facilitation, feedback, etc.)
        - Primary vs secondary CT outcomes
3. Exploring relationships within and between studies

   Patterns, similarities, and divergences in findings were explored across different stusdy types, AI applications, and CT measured. The synthesis was stratified where needed (e.g., primary vs. secondary CT outcomes).
4. Assessing robustness of the synthesis

   The credibility of findings was strengthened by linking results to study quality, triangulating findings from different study types (qualitative, mixed methods), and identifying areas of convergence and inconsistency.

3.9.2 Thematic Coding Procedures

To complement the narrative synthesis, a thematic analysis was conducted following a hybrid approach of deductive and inductive coding:

- Deductive phase: A coding framework was created based on the research questions and predefined themes (e.g., CT dimensions, AI function, assessment methods). This structure guided the initial coding of extracted data.
- Inductive phase: Emerging sub.themes and concepts not captured by the initial framework were identified during the line-by-line coding of qualitative findings and author interpretations.

The coding process was carried out in Excel by one reviewer and cross-validated by a second review. Themes were organized into a thematic coding matric (see Appendix C), which was used to structure the Results section (Section 4.5). Each theme was illustrated using representative study findings, and convergence across study types was noted when applicable.

Results

4.1 Study Selection

The study selection process is illustrated in the PRISMA 2020 flow diagram (Presented in the Appendix C). A total of 15,652 records were identified through database searching across Scopus (n=13,129), ERIC (n=1,663), and Web of Science (n=860). Following the removal of 14,695 records through database-applied filters and, then, 73 duplicates. 957 records remained for title and abstract screening.

Of these, 892 records were excluded based on the eligibility criteria. The full texts of 65 reports were retrieved and assessed for eligibility. A total of 43 reports were excluded at this stage for the following reasons: not an empirical study (n=3), retracted publication (n=1), no AI assessed (n=16), not assessing AI's impact on student's critical thinking (n=18), and not involving higher education students (n=5).

Ultimately, 22 studies met the inclusion criteria and were included in the review.

4.2 Study Characteristics

A total of 22 studies met the inclusion criteria (Figure 1) and were included in the systematic review These studies were published between 2022 and early 2025, with the majority appearing in 2023-2024, reflecting the recent surge of interest in critical thinking and AI's educational impact.

As seen in Table 2, the final selected studies exhibit diverse methodological designs and approaches to investigating the role of AI in assessing or fostering CT in educational contexts. Study designs were predominantly quasi-experimental (n=7), observational (n=4), cross-sectional (n=3), experimental (n=3), case study (n=3), action research (n=2), and design-based research (n=1). Mixed-methods approaches were the most common data type (n=15), followed by quantitative (n=5) and qualitative (n=2) designs. A wide range of AI tools were studied, with ChatGPT emerging as the most frequently investigated (n=13), followed by tools such as GPT-based chatbots, deepfake-generating GANs, text-to-speech technologies, and humanoid social robots.

In terms of CT as a learning objective, it was the primary outcome in 13 studies and a secondary outcome in 9. Measurement of CT was operationalized through various instruments, including questionnaires, surveys, rubrics, interviews, observational logs, and performance-based assessments. However, standardization varied greatly. While some studies employed well-validated rubrics or cited frameworks like the WPA Outcomes Statement, others relied on internally developed instruments or informal qualitative reflections.

Only 9 of the 22 studies provided a formal or scholarly definition of critical thinking, referencing established theoretical models such as those by Facione (1990), Paul and Elder, or the OECD. The remaining studies embedded CT implicitly through observed behaviors—such as reflection, argumentation, metacognition, and evidence evaluation—without formally defining the construct. Notably, even in the absence of explicit definitions, many studies detailed specific cognitive or metacognitive processes associated with CT.

Overall, the included studies reflect a growing interdisciplinary interest in leveraging AI technologies to engage students in higher-order thinking tasks. Despite methodological heterogeneity and inconsistency in the conceptual framing of CT, the collective findings suggest an emerging pattern: AI tools, when appropriately integrated, can support aspects of analytical reasoning, reflective thinking, and evaluative judgment—core components of critical thinking as understood in educational literature.

(Table of included studies)

Table 2. The table displays for each included study the citation, study design, ….and … . Reproduced from Barker et al.[191]

Note: The symbols ▯ and ✕ represent whether a clear CT definition was provided (Yes) or not (No), respectively.

4.4 Results of Individual Studies

4.5 Narrative Synthesis of Findings

The results of the included studies were synthesized narratively following a structured thematic analysis. A total of 22 studies were included, representing a variety of AI tools, educational contexts, and assessment approaches related to critical thinking (CT) in higher education. The findings were categorized into five major thematic domains, derived from the review's conceptual framework and supported by inductive coding: (1) definitions of CT, (2) assessment methods, (3) CT dimensions measured, (4) the role of AI tools, and (5) whether CT was a primary or secondary outcome.

4.5.1 Operational Definitions of Critical Thinking

Among the 22 included studies, only 8 provided an explicit, academically grounded definition of critical thinking, while the remaining 14 studies either embedded the concept implicitly or operationalized it through observable behaviors. Studies offering formal definitions (e.g., Studies 2, 5, 8, 9, 12, 15, 16, and 19) drew on established frameworks such as those by Facione, Lipman, and Paul & Elder. These definitions emphasized critical thinking as a core competency encompassing analysis, evaluation, synthesis, and self-regulation. Several of these studies took a multidimensional perspective, incorporating both cognitive and metacognitive dimensions, and linked critical thinking to broader educational frameworks and professional competencies. In contrast, studies without a formal definition (e.g., Studies 1, 3, 4, 6, 7, 10, 11, 13, 14, 17, 18, 20, 21, and 22) tended to infer critical thinking through learners' behaviors—such as questioning, analysis, or reflective writing—and measured it via rubrics, performance assessments, or qualitative indicators. These studies often emphasized domain-specific or context-driven manifestations of critical thinking, treating it as an emergent, practice-oriented skill. The divergence suggests two dominant conceptual camps: one viewing critical thinking as a transferable, theoretically defined cognitive construct, and the other adopting a pragmatic, context-sensitive approach grounded in observable learning outcomes. This bifurcation highlights a broader issue in the literature—namely, the lack of conceptual alignment across studies, which may limit comparability of results and complicate interpretation of AI's impact on critical thinking.

4.5.2 Aspects of CT Measured

The 22 studies reviewed assessed a broad and multidimensional range of critical thinking (CT) components. Despite methodological and contextual diversity, most studies aligned—implicitly or explicitly—with theoretical models such as Facione's Delphi Report, Ennis's taxonomy, and Halpern's framework. Thematic coding revealed five key CT dimensions.

Core cognitive-evaluative skills—including analysis, evaluation, synthesis, inference, and problem solving—were assessed in nearly all studies. Analysis was targeted in Studies 1, 7, 9, 11, and 22, requiring students to deconstruct information, identify elements, and distinguish authentic from AI-generated content. Evaluation appeared in Studies 1, 5, 7, 8, and 10, involving argument critique, source credibility assessment, and validation of AI outputs. Synthesis tasks in Studies 1, 10, and 16 asked students to integrate information across contexts to generate solutions, compare viewpoints, or design new approaches. Inference and problem-solving were central in Studies 8 and 22, where structured reasoning was used to address complex tasks requiring judgment and decision-making. These cognitive processes often formed the basis of structured assessments such as argument analysis, scenario-based tasks, and AI-feedback evaluations.

Metacognitive and reflective processes—including self-regulation, strategic reasoning, and reflective inquiry—were addressed in Studies 1, 2, 6, 7, 8, 12, 15, 19, and 20. These studies explored how learners monitored their reasoning, corrected misunderstandings, or evaluated their decision-making in real time. Reflective dialogue and Socratic questioning were used in Studies 2, 15, and 20 to stimulate metacognitive engagement, often mediated by AI prompts. This dimension highlights CT not only as a cognitive operation but also as an iterative, self-aware process.

Argumentation and communication featured prominently in several studies, with CT assessed through students' ability to structure, justify, and communicate reasoning. Study 1 focused on argumentative structure in written essays; Studies 3 and 6 evaluated coherence, clarity, and logical progression in student submissions. Study 13 analyzed legal case arguments, while

Studies 12 and 20 documented increased expressive confidence and fluency, emphasizing CT as both a reasoning and communication skill.

Epistemic judgment and information literacy emerged as particularly relevant in AI-supported learning environments. Studies 7–10 assessed the ability to detect bias, evaluate source reliability, and reconcile conflicting claims. In Study 10, students synthesized AI-generated summaries with scholarly content, applying critical literacy and epistemic vigilance. Ethical reflection was included in several studies, especially when interpreting AI feedback or addressing potential misinformation. This dimension reflects a growing need to integrate digital and epistemic competencies into CT instruction.

Behavioral, dispositional, and domain-specific competencies were addressed in contexts emphasizing practical application of CT. Study 4 assessed independent thinking and engagement as proxies for CT. Study 14 connected CT to overall academic performance, using learning outcomes as indicators of critical engagement. Studies 11, 13, 16, 18, and 22 embedded CT within applied tasks—such as design analysis, legal reasoning, and data interpretation—illustrating how CT manifests in real-world disciplinary settings as an observable, action-oriented skill.

In synthesis, CT across these studies is best understood through five interrelated dimensions: foundational cognitive-evaluative operations (analysis, evaluation, synthesis); metacognitive and reflective self-regulation; argumentation and communication of reasoning; epistemic and digital literacy for information credibility and ethical AI use; and domain-specific behavioral competencies evident in applied academic or professional contexts.

4.5.4 Role of AI Tools

The 22 studies included in this review employed a wide array of artificial intelligence (AI) tools to support or assess students' critical thinking (CT), ranging from conversational agents to assessment platforms and embodied AI. Five distinct functional categories emerged based on the role these tools played in CT development.

The most commonly used tools were **conversational AI and generative language models**, with 18 studies focusing on platforms like ChatGPT, Microsoft Copilot, or custom GPT-based chatbots. These tools were typically used for reflective questioning, argumentative writing, or facilitating dialogic engagement, positioning AI as both a content generator and cognitive partner in promoting analytical reasoning.

Feedback-driven and assessment-oriented AI, such as Studiosity or Bayesian tracking systems, featured in studies focused on personalized feedback, rubric-aligned evaluation, and gamified CT assessment. These tools provided adaptive learning support and insights into student progress, often operationalizing CT through structured, performance-based indicators.

In a third group, **content-generating AI tools**—including GANs, fake news bots, and platforms like YOU.com or Tome AI—were used to produce artifacts that students critically evaluated. This approach emphasized skills like source verification, bias detection, and epistemic judgment, treating AI outputs as catalysts for CT engagement rather than learning aids.

A smaller set of studies explored **assistive and embodied AI interfaces**, such as humanoid robots and text-to-speech systems, which aimed to foster reflective thinking, enhance accessibility, and support real-time interaction in educational tasks.

Finally, several studies investigated **domain-specific AI applications** in contexts like legal reasoning and engineering ethics, where AI was used to simulate case analysis, promote debate, or scaffold problem-solving in complex disciplinary tasks.

In synthesis, while the reviewed AI tools varied in function, interactivity, and domain specificity, a clear trend emerged: generative and dialogic tools were more frequently associated with higher-order CT engagement, while assessment and assistive tools tended to support structured, task-bound outcomes. Importantly, not all studies aligned AI integration intentionally with CT learning objectives, highlighting variability in design quality and pedagogical coherence.

4.6 Reported Impact of AI on Students' Critical Thinking

Across 22 included studies, the impact of AI on students' critical thinking was found to vary substantially depending on context, type of AI intervention, and the extent of human facilitation. Interventions included generative tools (e.g., ChatGPT), intelligent tutoring systems, social robotics, and text-to-speech technologies, applied across disciplines such as programming, literature, legal education, and medical informatics. Designs ranged from controlled experiments to self-reported surveys, with both quantitative and qualitative data.

Findings suggest a continuum of influence. When AI was integrated to support, rather than replace, cognitive engagement—through features such as feedback, counterarguments, or reflective questioning—it enhanced critical thinking outcomes. Studies consistently reported statistically significant gains in analytical skills, argumentation, and metacognitive engagement. For example, one study showed critical thinking scores increased from 3.40 to 8.21 (η^2 = 0.63–0.69), and others documented improvements in debate performance, literature analysis, and domain-specific reasoning in programming.

AI's effectiveness was attributed to its ability to provide personalized, immediate feedback and serve as a nonjudgmental partner for inquiry. In self-directed settings, students engaged with AI to test assumptions and iterate ideas, particularly in contexts with limited access to human feedback. Notably, students from non-technical backgrounds reported enhanced confidence and self-directed reasoning when using AI in data analytics contexts.

However, these benefits were contingent on thoughtful integration. Studies comparing AI-only environments with human-guided or hybrid models found that exclusive reliance on AI often led to shallow learning, cognitive offloading, and reduced critical reflection. Hybrid models, combining AI with instructor support, consistently produced stronger outcomes by maintaining depth of inquiry and mitigating overreliance.

Some studies also identified risks. Overuse of AI tools was associated with lower test performance and academic integrity concerns, with students occasionally bypassing analytical engagement in favor of AI-generated outputs. Age, cultural background, and educational context moderated these effects, with older students and less rote-based systems showing greater discernment.

The influence of AI varied across domains. In mathematics and programming, AI enhanced conceptual understanding when integrated with active guidance. Students' perceptions of AI's usefulness and usability positively correlated with improved critical thinking outcomes (e.g., β =

0.215 and 0.205, respectively). One study reported that 64% of participants noted substantial improvements in their critical thinking after engaging with generative AI tools.

In synthesis, AI appears most effective when employed as a cognitive enhancer within structured, reflective learning environments. It supports iterative reasoning, problem-solving, and evaluation when coupled with pedagogical design and instructor feedback. Conversely, using AI as a substitute for cognitive effort can diminish learning depth.

This review underscores the need for curriculum designs that emphasize AI literacy and critical evaluation of machine-generated outputs. Standardizing outcome measures and conducting longitudinal research will be essential to determine the durability of AI-supported critical thinking gains. Ensuring ethical, balanced integration remains key to optimizing AI's role in education.

4.7 Subgroup Trends and Patterns

Two key moderating factors emerged in the synthesis: the disciplinary context of learning activities and the extent of instructor involvement alongside AI tools. The disciplinary context shaped how AI influenced critical thinking across different fields. In STEM disciplines such as computer science and engineering, AI was typically used to support problem-solving and logical reasoning. Students working with tools like coding assistants or problem generators showed gains in tasks like debugging, modeling, and applying formal logic, skills aligned with critical thinking in technical domains.

In contrast, in the humanities and social sciences, AI tools were used to stimulate debate, expose bias, or analyze texts. These applications emphasized critical thinking through argument quality, multi-perspective analysis, and ethical reasoning. For example, students in literature classes using AI text analyzers engaged more deeply in narrative critique, while law students simulated courtroom arguments with AI and critically evaluated their validity, developing skills such as identifying fallacies and weighing evidence. In language and communication contexts, such as academic writing or reading comprehension, students used AI to generate or refine content, improving their ability to identify inaccuracies, structure arguments, and enhance clarity. Those who edited AI-generated drafts, for instance, demonstrated more critical engagement than peers who wrote without that intermediate step. These findings suggest that aligning AI use with the cognitive demands of each discipline enhances its effectiveness, making critical thinking activities more relevant and integrated into existing pedagogical frameworks.

Instructor involvement also played a major role in determining AI's impact on critical thinking. Studies that employed hybrid or teacher-guided models, where educators directed how AI tools were used, framed tasks, or prompted reflection, tended to report stronger critical thinking outcomes. For instance, students participating in instructor-led debates supplemented by AI-generated arguments engaged in deeper analysis than those who worked independently with AI. Teacher presence helped mitigate risks such as overreliance on AI or superficial reasoning, as educators could guide interpretation, raise questions, and encourage skepticism. Conversely, in studies where AI was used autonomously without significant instructor input, results were often weaker or neutral. In such cases, students frequently accepted AI outputs at face value, bypassing deeper analysis. Some studies noted that students working alone with chatbots tended to disengage from critical evaluation, illustrating a tendency toward cognitive offloading in the absence of teacher scaffolding.

The evidence underscores that AI is most effective for fostering critical thinking when combined with active instructor facilitation. Educators who structure AI experiences with follow-up questions, critical reflection prompts, and justification tasks help students engage more thoughtfully with AI content. Without this guidance, students may not challenge the technology or extend their reasoning beyond initial outputs. Overall, the review highlights that both disciplinary alignment and human facilitation are essential for maximizing AI's potential to support critical thinking in higher education.

Discussion

5.1 Principal Findings

This systematic review synthesized findings from 22 studies conducted between 2022 and 2025, each examining the role of artificial intelligence (AI) tools in shaping critical thinking (CT) among higher education students. The review followed the **PRISMA 2020 guidelines** and adopted a narrative synthesis approach due to the heterogeneity of interventions, designs, and outcome measures.

Key findings indicate that the impact of AI on CT is **positive in most contexts**, particularly when AI is implemented as a **scaffold for cognitive engagement** rather than a **replacement for analytical effort**. The most effective AI tools were those that promoted:

- Reflective dialogue (e.g., via ChatGPT)
- Real-time feedback and error correction (e.g., via intelligent tutoring systems or rubrics)
- Evaluation of misinformation or AI-generated content (e.g., fake news bots or GANs)

Notably, **63.6% of studies treated CT as a primary outcome**, employing formal definitions and intentional measurement strategies, while the remaining **36.4% approached CT indirectly**, embedding it within broader academic or problem-solving tasks. Studies that explicitly defined CT and used validated instruments tended to report **stronger learning gains** and more structured analytical outcomes.

The overall synthesis reveals that **AI tools can enhance critical thinking**, but only under certain conditions—namely when instructional design, tool functionality, and human facilitation are closely aligned with CT development goals.

5.2 Strengths and Limitations of the Included Studies

The studies included in this review provide meaningful insights into the role of AI in fostering critical thinking within higher education, though several limitations should be noted when interpreting their findings.

Among the strengths, a significant majority (77%) employed **mixed methods** or **quasi-experimental designs**, offering both quantitative evidence and qualitative perspectives on student reasoning processes. The studies also reflected **disciplinary diversity**, spanning fields such as law, computer science, and literature, thereby illustrating the applicability of AI tools in supporting critical thinking across varied curricular and cognitive contexts. Additionally, many studies drew on **multiple data sources** enhancing internal validity through **triangulation**.

However, limitations were apparent. There was **substantial variation** in how critical thinking was defined and assessed: only 8 studies explicitly provided a definition of CT, and just one employed a dedicated critical thinking assessment instrument. Moreover, a reliance on **self-reported outcomes** or student perceptions was common, raising concerns about the validity of claims regarding actual skill development. Finally, many studies utilized **convenience sampling** or involved **small sample sizes**, which constrains the generalizability of their findings.

5.3 Strengths and Limitations of the Review Process

This review exhibited several methodological strengths that enhance its transparency and rigor. It followed the **PRISMA 2020 guidelines**, ensuring reproducibility and a structured approach throughout all review stages. Searches were conducted in three leading academic databases—**Scopus**, **Web of Science**, and **ERIC**—targeting peer-reviewed, high-quality studies in education and educational technology. A carefully developed and database-specific search strategy combined keywords related to **AI**, **critical thinking**, and **higher education**.

Screening and selection followed a **two-stage process**, with independent reviewer assessments and consensus-based conflict resolution, reinforcing the reliability of inclusion decisions.

Nonetheless, certain limitations must be acknowledged. The review included only **English-language studies**, introducing potential **language bias**, and excluded **grey literature** (e.g., conference papers, preprints), which may have led to **publication bias**. The **diversity of study designs and critical thinking definitions** prevented meta-analysis, limiting the ability to quantify effect sizes or statistically generalize findings.

Another notable limitation was the **overrepresentation of ChatGPT-focused studies**—18 of the 22 included articles centered on this tool. While this reflects ChatGPT's rapid adoption since its late-2022 release, it introduces **tool-specific bias**, potentially skewing how critical thinking is conceptualized within ChatGPT-mediated learning contexts. This concentration likely stems from the review's **2022–2025 inclusion window**, designed to capture recent developments but at the expense of earlier studies involving alternative AI tools. Future reviews may benefit from broader timeframes or comparative frameworks that assess definitions and outcomes across diverse AI applications.

Despite these constraints, the review maintained a high level of methodological transparency and adhered closely to best practices in **qualitative evidence synthesis**, offering a valuable contribution to understanding how critical thinking is framed in AI-enhanced higher education research.

5.4 Comparison with Prior Reviews or Frameworks

Few systematic reviews have specifically addressed the intersection of artificial intelligence and critical thinking in higher education. Most existing reviews tend to focus on AI's role in learning analytics, personalized instruction, or overall academic performance, rather than examining critical thinking as a distinct outcome. This review supports established educational theories by showing that classic definitions of critical thinking (such as those by Facione, Ennis, or Paul and Elder) remain relevant even in AI-supported learning contexts. Students still need to analyze arguments, reflect on information, and make reasoned judgments, whether or not AI tools are involved. However, the review also reveals that many studies fail to apply these frameworks rigorously, leading to inconsistent definitions, measurements, and reporting of critical thinking, an issue long noted in the literature.

This review contributes new insights by identifying AI-specific factors that influence critical thinking, which traditional frameworks did not account for. For example, it distinguishes between AI tools that generate content and those that provide feedback, noting that generative AI may encourage open-ended inquiry while feedback-oriented AI supports more structured problem-solving. These nuances refine previous research in educational technology by clarifying how different AI functions align with different dimensions of critical thinking.

In contrast to broader reviews on educational technology that emphasize engagement or general achievement, this synthesis offers a more targeted, evidence-based examination of how AI tools can support, or fail to support, critical thinking processes. It confirms that technology's effect on critical thinking is conditional, depending on how the tool is used and in what context. This finding aligns with constructivist theories that emphasize the role of active learning design. Ultimately, the review reinforces the idea that merely introducing AI into educational settings is not sufficient; its integration must be thoughtfully designed to support the specific cognitive processes that critical thinking entails.

5.5 Implications for Educational Practice and Policy

The findings of this review highlight several key implications for educational practice and policy. AI should be implemented as a cognitive scaffold that supports, rather than replaces, critical thinking. When used to prompt reflection, challenge assumptions, or guide revisions, AI tools can enhance students' cognitive engagement. However, instructor facilitation remains crucial, as human educators offer ethical oversight, nuanced feedback, and individualized support that AI alone cannot provide. The consistent success of hybrid models underscores the importance of integrating AI with active human guidance. To prepare students for AI-enhanced learning environments, AI literacy must be embedded in the curriculum, equipping learners not only with operational skills but also with the ability to critically evaluate AI outputs, recognize bias, and understand tool limitations. Assessment frameworks must also evolve to include validated instruments that reflect both disciplinary goals and the capabilities of AI-supported learning. Lastly, ethical guidelines and academic integrity policies require updating to address risks such as cognitive offloading, plagiarism, and inappropriate use of AI-generated content in critical thinking tasks.

5.6 Recommendations for Future Research

Future research should adopt theory-based and consistently defined concepts of critical thinking, aligning them with instructional design and assessment practices. The use of validated instruments is necessary to ensure measurement reliability and minimize self-report bias. Longitudinal studies are needed to evaluate the durability, transferability, and practical application of AI-supported critical thinking beyond immediate learning contexts.

Further exploration is required in underrepresented disciplines such as health sciences, visual arts, and social work, where critical thinking is vital but underexplored in AI-mediated environments. Cultural, demographic, and contextual variables should be systematically examined to assess their influence on equitable AI adoption.

Comparative research should investigate the effectiveness of AI-only versus hybrid instructional models to identify optimal strategies for critical thinking development. Future assessment frameworks should also incorporate ethics and metacognition to fully capture students' reflective and evaluative capacities within AI-enhanced settings.

Key gaps include the absence of research synthesizing AI's role among secondary and non-student adult learners, a lack of standardized classification of AI tools, and no review tracing the historical evolution of these technologies in education. Additionally, while there is a large body of work on critical thinking assessment, no existing review spans all methodological approaches across disciplines. Addressing these gaps would support both empirical inquiry and the development of more comprehensive, integrative reviews.

Conclusion

6.1 Summary of Key Insights

This systematic review synthesized 22 empirical studies published between 2022 and 2025 that investigated the impact of artificial intelligence (AI) tools on critical thinking (CT) development in higher education. Drawing from diverse academic contexts, AI modalities, and study designs, the review aimed to clarify how critical thinking is defined, measured, and affected by AI-enhanced learning interventions.

The evidence confirms that AI technologies hold strong potential as cognitive amplifiers—tools that, when embedded thoughtfully into educational design, can foster students' engagement in analytical reasoning, argumentation, evaluation, and reflective inquiry. However, this potential is not universally realized across the literature. The impact of AI on CT appears to vary depending on:

- The type and interactivity of the AI tool (e.g., ChatGPT, intelligent tutoring systems, generative AI platforms)
- The disciplinary context and cognitive demands of the learning tasks
- The presence or absence of human instructor facilitation
- The conceptual clarity and measurement specificity of CT within each study

Only a minority of studies (8 of 22) offered a formal, theory-grounded definition of critical thinking, and even fewer employed dedicated CT measurement instruments. Instead, many relied on rubric-based assessments, student self-perception surveys, or domain-specific performance tasks. This definitional and methodological heterogeneity complicates cross-study comparisons and underscores a need for greater conceptual alignment in future research.

Nonetheless, consistent patterns emerged. Studies that integrated AI as a complement to human-led instruction, and that emphasized active engagement, questioning, and reflection, tended to report positive effects on CT outcomes. In contrast, studies that positioned AI as a standalone cognitive substitute—especially those relying heavily on passive student-AI interactions—reported neutral or even negative effects, often linked to cognitive offloading or superficial learning.

The review also revealed that critical thinking is typically conceptualized across five major dimensions in AI-enhanced learning environments:

1. Core cognitive-evaluative skills (analysis, inference, evaluation, synthesis)
2. Metacognitive and reflective processes
3. Argumentation and communication
4. Epistemic judgment and information literacy
5. Behavioral and domain-specific applications

These categories, grounded in frameworks by Facione, Ennis, and Halpern, offer a structured lens for future study design, assessment development, and AI integration in pedagogy.

6.2 Final Reflections

In a time of rapidly evolving educational technologies, this review underscores the gap between the growing use of AI in higher education and the lack of clear theoretical and empirical consensus on its impact on critical thinking. While there is promising evidence that AI can support the development of critical thinking, its effectiveness depends on how it is implemented, what aspects of critical thinking are being measured, and how the concept itself is defined. The review calls for future research to be more rigorous by grounding critical thinking frameworks in established theory, aligning assessment tools accordingly, and avoiding vague claims of improvement without specifying which skills are being developed. It also emphasizes the importance of combining AI tools with reflective, human-guided learning strategies. As AI becomes more prevalent in classrooms, the authors stress the need for curricula that foster AI literacy, ethical use, and metacognitive skills. Teaching students not just to use AI, but to question and evaluate it critically, is essential for ensuring it enhances rather than undermines their ability to think independently. Overall, the review serves as a starting point for clearer conceptual understanding and responsible innovation in AI-supported education.

References

1. **Al-Othman, A.A.M.** (2023) 'Assistive technology for enhancing Saudi female undergraduate students' reading cognitive abilities: An investigation of reading models and their impacts on reading performance', *Journal of Language Teaching and Research*, 15(1), pp. 116–127. doi:10.17507/jltr.1501.14.
2. **Dai, Y.** et al. (2023) 'CHATGPT and its impact on research supervision: Insights from Australian Postgraduate Research Students', *Australasian Journal of Educational Technology*, 39(4), pp. 74–88. doi:10.14742/ajet.8843.
3. **Dasari, D.** et al. (2024) 'Chatgpt in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning', *Frontiers in Education*, 8. doi:10.3389/feduc.2023.1295413.
4. **Eltahir, M.E.** and **Mohd Elmagzoub Babiker, F.** (2024) 'The influence of artificial intelligence tools on student performance in e-learning environments: Case study', *Electronic Journal of e-Learning*, 22(9), pp. 91–110. doi:10.34190/ejel.22.9.3639.
5. **Fakour, H.**, & **Imani, M.** (2025). Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills. *Frontiers in Education*, 10. https://doi.org/10.3389/feduc.2025.1528603
6. **Ganjoo, R.** et al. (2024) 'Beyond boundaries: Exploring a generative artificial intelligence assignment in graduate, online science courses', *Journal of Microbiology & Biology Education*, 25(3). doi:10.1128/jmbe.00127-24.
7. **Huang, X.** (2024) 'Research on the construction of English intelligent teaching mode in colleges and universities facilitated by Artificial Intelligence Technology', *Applied Mathematics and Nonlinear Sciences*, 9(1). doi:10.2478/amns-2024-0657.
8. **Jayasinghe, S.** (2024) 'Promoting active learning with CHATGPT: A constructivist approach in Sri Lankan Higher Education', *Journal of Applied Learning & Teaching*, 7(2). doi:10.37074/jalt.2024.7.2.26.
9. **Kim, D.**, **Majdara, A.**, & **Olson, W.** (2024). A Pilot Study Inquiring into the Impact of ChatGPT on Lab Report Writing in Introductory Engineering Labs. *International Journal of Technology in Education.* https://doi.org/10.46328/ijte.691
10. **López-Caudana, E.**, **George Reyes, C. E.**, & **Avello Martínez, R.** (2024). Developing the skills for complex thinking research: a case study using social robotics to produce scientific papers. *Frontiers in Education*, 9. https://doi.org/10.3389/feduc.2024.1322727
11. **Magalhães Araujo, S.** and **Cruz-Correia, R.** (2024) 'Incorporating chatgpt in Medical Informatics Education: Mixed methods study on student perceptions and experiential integration proposals', *JMIR Medical Education*, 10. doi:10.2196/51151.
12. **Melisa, R.**, **Rahyuni, M.**, **Ashadi, A.**, **Triastuti, A.**, **Hidayati, S.**, **Salido, A.**, **Luansi Ero, P. E.**, **Marlini, C.**, **Zefrin, Z.**, & **Al Fuad, Z.** (2025). Critical thinking in the age of AI: A

systematic review of AI’s effects on higher education. *Educational Process: International Journal*, 14(1). https://doi.org/10.22521/edupij.2025.14.31

13. **Michalon, B.** and **Camacho-Zuñiga, C.** (2023) ‘CHATGPT, a brand-new tool to strengthen timeless competencies’, *Frontiers in Education*, 8. doi:10.3389/feduc.2023.1251163.
14. **Musi, E.** et al. (2023) ‘Developing misinformation immunity: How to reason-check fallacious news in a human–computer interaction environment’, *Social Media + Society*, 9(1). doi:10.1177/20563051221150407.
15. **Murillo-Ligorred, V.** et al. (2023) ‘Knowledge, integration and scope of Deepfakes in arts education: The development of critical thinking in postgraduate students in primary education and master’s degree in Secondary Education’, *Education Sciences*, 13(11), p. 1073. doi:10.3390/educsci13111073.
16. **Mustafa, M. Y.**, **Tlili, A.**, **Lampropoulos, G.**, **Huang, R.**, **Jandrić, P.**, **Zhao, J.**, **Salha, S.**, **Xu, L.**, **Panda, S.**, **Kinshuk**, **López‑Pernas, S.**, & **Saqr, M.** (2024). A systematic review of literature reviews on artificial intelligence in education (AIED): A roadmap to a future research agenda. *Smart Learning Environments*, 11(1), Article 59. https://doi.org/10.1186/s40561-024-00350-5
17. **Naatonis, R.N.** et al. (2024) ‘Evaluation of problem based Gamification Learning (PBGL) model on critical thinking ability with artificial intelligence approach integrated with CHATGPT API: An experimental study’, *Qubahan Academic Journal*, 4(3), pp. 485–520. doi:10.48161/qaj.v4n3a919.
18. **Page, M. J.**, **McKenzie, J. E.**, **Bossuyt, P. M.**, et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71
19. **Qawqzeh, Y.K.** (2024) ‘Exploring the influence of student interaction with CHATGPT on critical thinking, problem solving, and creativity’, *International Journal of Information and Education Technology*, 14(4), pp. 596–601. doi:10.18178/ijiet.2024.14.4.2082.
20. **Rai, D. H.** (2024). Artificial intelligence through time: A comprehensive historical review [Preprint]. *ResearchGate*. https://doi.org/10.13140/RG.2.2.22835.03364
21. **Robillos, R.** (2024) ‘Synergizing generative pre-trained transformer (GPT) chatbots in a process-based writing paradigm to enhance university students’ writing skills’, *Journal of Language and Education*, 10(3), pp. 79–94. doi:10.17323/jle.2024.18708
22. **Ruiz-Rojas, L.I.**, **Salvador-Ullauri, L.** and **Acosta-Vargas, P.** (2024) ‘Collaborative working and critical thinking: Adoption of generative artificial intelligence tools in higher education’, *Sustainability*, 16(13), p. 5367. doi:10.3390/su16135367.
23. **Sasikala, P.**, & **Ravichandran, R.** (2024). Study on the impact of artificial intelligence on student learning outcomes. *Journal of Digital Learning and Education*, 4(2), 145–155. https://doi.org/10.52562/jdle.v4i2.1234

24. **Tsai, Y.-C.** (2024) ‘Empowering students through active learning in Educational Big Data Analytics’, *Smart Learning Environments*, 11(1). doi:10.1186/s40561-024-00300-1.
25. **Walter, Y.** (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(15). https://doi.org/10.1186/s41239-024-00448-3
26. **Wang, W.**, **Xu, Zhilang** and **Xu, Zichun** (2024) ‘Changes and challenges of legal education in the era of Generative Artificial Intelligence: Chinese experience’, *Journal of Infrastructure, Policy and Development*, 8(8), p. 5600. doi:10.24294/jipd.v8i8.5600.
27. **Yücel, A.** (2025). Critical thinking and education: A bibliometric mapping of the research literature (2005–2024). *Participatory Educational Research*, 12, 137–163. https://doi.org/10.17275/per.25.23.12.2
28. **Zhou, X.**, **Teng, D.** and **Al-Samarraie, H.** (2024) ‘The mediating role of Generative AI self-regulation on students’ critical thinking and problem-solving’, *Education Sciences*, 14(12), p. 1302. doi:10.3390/educsci14121302.

Appendices

Appendix A. Full Search Strategies

- A-1. Database: ERIC

  Date searched: February 23, 2025

  Search fields used: TITLE-ABS-KEY

  Full search string:

  TITLE-ABS-KEY ("Critical Thinking") AND ("Higher Education" OR "University" OR "College" OR "Undergraduate*") AND ("Assess*" OR "Evaluat*" OR "Measure*" OR "Rubric*" OR "Framework*" OR "Instrument*" OR "Tool*" OR "Test*") AND ("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning" OR "Generative AI" OR "Large Language Models" OR "LLM" OR "Natural Language Processing" OR "NLP" OR "Intelligent Tutoring Systems" OR "ITS" OR "AI Chatbot*" OR "Virtual Assistant*" OR "AI-Based Feedback" OR "AI-Powered Feedback" OR "AI-Assisted Feedback" OR "AI-Assisted Learning" OR "Automated Feedback" OR "Adaptive Learning AI" OR "Gamifi*")

  Filters applied:

  - Publication date: Since 2021 (last 5 years)
  - Descriptor: Critical Thinking
  - Publication type: Journal Articles, Reports - Research
  - Education Level: Higher Education, Postsecondary Education
  - Peer reviewed only: Yes
  - Full text available on ERIC: Yes

  Total results retrieved: 91

- A-2. Database: SCOPUS

  Date searched: February 23, 2025

  Search fields used: TITLE-ABS-KEY

  Full search string:

TITLE-ABS-KEY ("Critical Thinking") AND ("Higher Education" OR "University" OR "College" OR "Undergraduate*") AND ("Assess*" OR "Evaluat*" OR "Measure*" OR "Rubric*" OR "Framework*" OR "Instrument*" OR "Tool*" OR "Test*") AND ("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning" OR "Generative AI" OR "Large Language Models" OR "LLM" OR "Natural Language Processing" OR "NLP" OR "Intelligent Tutoring Systems" OR "ITS" OR "AI Chatbot*" OR "Virtual Assistant*" OR "AI-Based Feedback" OR "AI-Powered Feedback" OR "AI-Assisted Feedback" OR "AI-Assisted Learning" OR "Automated Feedback" OR "Adaptive Learning AI" OR "Gamifi*")

Filters applied:

- Publication years: 2022 - 2025
- Subject area: Social Sciences, Computer Science, Psychology, Multidisciplinary
- Language: English
- Document type: Article
- Open access status: Gold
- Keyword exclusions: Primary Education, Secondary Education, Systematic Review, Middle Aged, Teacher, Teachers, Systematic Literature Review, Adolescent

Total results retrieved: 822

- A-3. Database: WoS

Date searched: February 23, 2025

Search fields used: TS

Full search string:

TS=("Critical Thinking")

AND TS=("Higher Education" OR "University" OR "College" OR "Undergraduate*")

AND TS=("Assess*" OR "Evaluat*" OR "Measure*" OR "Rubric*" OR "Framework*" OR "Instrument*" OR "Tool*" OR "Test*")

AND TS=("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning" OR "Generative AI" OR

"Large Language Models" OR "LLM" OR "Natural Language Processing" OR "NLP" OR "Intelligent Tutoring Systems" OR "ITS" OR "AI Chatbot*" OR "Virtual Assistant*" OR "AI-Based Feedback" OR "AI-Powered Feedback" OR "AI-Assisted Feedback" OR "AI-Assisted Learning" OR "Automated Feedback" OR "Adaptive Learning AI" OR "Gamifi*")

Filters applied:

- Publication date: Last 5 years
- Publication years: 2022, 2023, 2024, 2025
- Web of Science categories: Education & Educational Research; Psychology, Educational; Education, Scientific Disciplines; Computer Science, Artificial Intelligence; Multidisciplinary Sciences
- Language: English
- Document type: Article, Early access
- Open access status: Gold, Open Access

Total results retrieved: 117

Appendix B. Thematic Coding Matrix

Studies	Study Design	Data Type	AI Tools Studied	CT as an Outcome	Data Collection Assessment Instruments for CT	Aspects of Critical Thinking Measured:		CT Definition	
(Murillo-Ligorred et al., 2023)	Quasi-Experimental	Qualitative	Deepfake images (created using Generative Adversarial Networks - GANs)	Primary	Questionnaire	-Distinguishing real vs. manipulated images -Argumentation quality in written responses -Recognition of ethical implications -Visual/media literacy and manipulation evaluation -Reflective dialogue engagement -Contextual and practical knowledge integration	[Orig]	NO	No formal definition; CT is implied through inquiry, integration of visual/digital literacy, ethical reflection, and social awareness within arts education.

(Fakour & Imani, 2025)	Observational	Mixed Methods	ChatGPT	Primary	Survey Interview	-Analytical Thinking -Open-ended Questioning -Reflective Thinking -Comparative Evaluation	[Orig]	YES	CT defined as involving analysis, evaluation, synthesis, self-reflection, and open-ended inquiry. Grounded in Socratic philosophy and OECD frameworks.
(Eltahir & Mohd Elmagzoub Babiker, 2024)	Quasi-Experimental	Mixed Methods	Kahoot! Chat GPT Studiosity	Secondary	Rubric-based assessment	-Understanding instructional design models (ADDIE, ASSURE) -Evaluation of model strengths and weaknesses -Creativity in model development -Integration of feedback to improve clarity and coherence	[Sum]	NO	No formal definition; CT operationalized via rubric assessing analysis, creativity, feedback integration, and argumentative clarity in instructional design essays.

						-Quality of academic writing -Persuasiveness and overall argumentative insight			
(Huang, 2024)	Experimental	Quantitative	Bayesian knowledge-tracking models (BKT and BF-BKT)	Secondary	(No specific instrument provided for critical thinking)	-Independent thinking inferred from student participation -Engagement measured through classroom behavior observation -Learning mastery tracked via Bayesian knowledge-tracing models	[Inf]	NO	CT treated as a subcomponent of learning ability; no definition given, but measured via student performance scores in an AI-supported English class.
(Zhou et al., 2024)	Cross-Sectional Correlational	Quantitative	Generative AI tools	Primary	Questionnaire	-Analysis, evaluation, and reflection -Questioning the validity of AI-generated	[Sum]	YES	CT defined as “the art of evaluating cognitive processes,” encompassing

						ideas -Focus on evidence-based judgment rather than synthesis			analysis, reasoning, open-mindedness, and metacognition (Paul & Elder; Davis & Barnett).
(Robillos, 2024)	Quasi-Experimental	Mixed Methods	GPT-based chatbots	Secondary	Rubric Pre-Test and Post-Test Quiz Interview Descriptive Checklist	-Argument formulation and logical structure in writing -Metacognitive reflection and goal setting -Monitoring comprehension of chatbot responses -Identifying and correcting writing errors -Self-regulated learning and error awareness	[Sum]	NO	No formal definition; CT demonstrated through reflection, metacognition, self-monitoring, and error correction during chatbot-supported writing tasks.

(Magalhães Araujo & Cruz-Correia, 2024)	Observational	Mixed Methods	ChatGPT	Secondary	Questionnaire	-Evaluating credibility and bias in AI outputs -Decision-making and reflection in applying AI responses -Understanding and explaining complex systems -Reformulating queries based on feedback -Comparing perspectives for cross-verification	[Sum]	NO	No formal definition; CT is embedded in student activities like evaluating ChatGPT content, rephrasing queries, and reflecting on AI limitations.
(Michalon & Camacho-Zuñiga, 2023)	Action Research	Mixed Methods	ChatGPT	Secondary	Observation Survey	-Identifying AI-generated inaccuracies -Self-correction and skepticism toward ChatGPT -Contrasting known vs. generated	[Sum]	YES	CT defined through two sources: identifying reasoning flaws (Olivares et al., 2021) and as “responsible,

						information -Reflecting on decision-making accuracy			criteria-based, self-correcting, context-sensitive thinking” (Lipman, 1988).
(López-Caudana et al., 2024)	Case study Quasi-Experimental	Mixed Methods	NAO humanoid social robots	Primary	Questionnaire (pre/post) Rubric	-Analyzing arguments and detecting false claims -Identifying research problems -Evaluating evidence-based solutions	[Orig]	YES	CT described as part of "complex thinking," involving analysis, synthesis, evaluation, and evidence-based reasoning (Ramírez-Montoya & Sanabria).
(Costa et al., 2024)	Experimental Correlational Cross-Sectional	Mixed Methods	ChatGPT	Secondary	Questionnaire	-Verifying AI-generated ideas with prior knowledge -Evaluating credibility and identifying misinformation	[Sum]	NO	No explicit definition; CT viewed as a learning goal and practiced through verification, analysis, and

						-Expanding or revising content using bibliographic evidence -Synthesizing improved output for final submission			enrichment of AI-generated academic content.
(Kim et al., 2024)	Quasi-Experimental	Mixed Methods	ChatGPT	Primary	Rubric Focus Group	-Analyzing and interpreting lab data -Linking evidence to conclusions -Evaluating relevance and technical accuracy	[Sum]	NO	No formal definition; CT assessed via rubric focusing on data interpretation, evidence-claim connection, and evaluation of lab report revisions.
(Musi et al., 2023)	Quasi-Experimental	Mixed Methods	Fake News Immunity Chatbot Vaccinating News Chatbot	Primary	Questionnaire Survey	-Reflection -Insights -Focus -Argumentation -Explanation -Assessing Facts and Evidence -Distinguishing	[Orig]	YES	CT defined by Facione (1990) as purposeful, self-regulatory judgment involving interpretation, analysis,

						-Changing Assumptions			evaluation, inference, and explanation.
(Wang et al., 2024)	Case study	Mixed Methods	ChatGPT Wenxin Yiyan LLaMa Palm	Primary	Observation logs	-Legal reasoning and argumentation development -Evaluating legal facts and principles -Interpreting multiple legal perspectives -Applying reasoning in cross-cultural case discussions	[Sum]	NO	No formal definition; CT described as a skill set involving questioning, reasoning, independent judgment, and multi-perspective legal analysis.
(Dasari et al., 2024)	Quasi-Experimental	Mixed Methods	ChatGPT	Secondary	Posttest math performance (quantitative proxy) Interview Classroom observations	-Inference-based indicators of reflection and reasoning -Classroom comparison suggests higher-order thinking when guided -Critical	[Inf]	NO	No definition provided; CT inferred from students' ability to reflect, ask deep questions, and apply dialectical reasoning—lar

						questioning inferred from instructor-led interactions -ChatGPT-only learning shown to reduce reflective engagement			gely absent in AI-only learning.
(Dai et al., 2023)	Case study	Qualitative	ChatGPT	Primary	Interview	-Challenging assumptions using devil's advocate strategies -Considering multiple perspectives via counterarguments -Reflective critique and bias identification -Strategic questioning and iterative self-revision -Verifying AI content against	[Sum]	YES	CT defined as the development of intellectual rigor—thinking in new ways, analyzing, questioning, and recognizing flaws in arguments (Lee, 2008).

						reliable sources			
(Ruiz-Rojas et al., 2024)	Cross-Sectional	Mixed Methods	Canva Chat PDF YOU.COM ChatGPT Tome AI Google Docs Zoom	Primary	Survey	-Evaluation and synthesis of information -Generation of innovative ideas and solutions	[Orig]	YES	CT defined as reflective analysis, synthesis, questioning assumptions, examining evidence, and drawing informed conclusions (Thornhill-Miller et al., 2023).
(Qawqzeh, 2024)	Cross-Sectional	Mixed Methods	ChatGPT	Primary	(Survey) Questionnaire	Not specified; students only reported perceived improvement in CT levels	[Inf]	NO	No explicit definition; CT understood through self-reported skills like analysis, argument evaluation, and recognition of bias in AI-assisted

									learning.
(Al-Othman, 2023)	Quasi-Experimental	Quantitative	Text-to-Speech Technology (TTS)	Primary	Reading Comprehension Test (Pretest & Posttest) Questionnaire #1 Questionnaire #2	-Critical Reading (QA): Reading with a critical perspective (e.g., identifying author intent, implied meaning) -Problem-Solving (QB): Solving problems and drawing conclusions based on the reading text	[Orig]	NO	No formal definition; CT emphasized as essential for comprehension and problem-solving in reading. Measured via TTS-supported reading assessments.
(Naatonis et al., 2024)	Quasi-Experimental	Quantitative	ChatGPT API	Primary	Pre-test and post-test assessments Critical thinking skills instrument	-Analytical reasoning -Logical application -Systematic problem-solving -Sequential structured thinking, -Adaptability through reflective	[Orig]	YES	CT defined as analyzing and constructing arguments based on logic and evidence (Kuhn, 2019). Highlights CT's role in decision-making and combating

						feedback			misinformation.
(Jayasinghe, 2024)	Action research	Qualitative	ChatGPT	Secondary	Interview	-Problem-solving and analysis of real-world issues -Synthesis and application of multi-source information -Reflective thinking and self-assessment -Questioning assumptions and evaluating alternatives -Exposure to multiple viewpoints and argumentative strategies -Confidence in expressing complex ideas	[Sum]	NO	No definition provided; CT implied as an evolving skill developed through active learning, questioning, and self-reflection in a constructivist framework.

(Ganjoo et al., 2024)	Quasi-Experimental	Mixed Methods	ChatGPT Microsoft Copilot	Primary	Rubric Survey	-Evaluation and integration of AI-generated and peer-reviewed content	[Orig]	**NO**	No formal definition; CT inferred from tasks requiring evaluation of AI outputs, integration with scholarly sources, and synthesis of information.
(Tsai, 2024)	Design-Based Research	Mixed Methods	ChatGPT API	Primary	Rubric Self/peer assessment Process observations (live coding/debugging sessions)	-Problem-solving with real-world datasets -Thematic reasoning and pattern identification -Interpretive analysis and inquiry-based learning -Judgment refinement through iterative questioning -Evaluating assumptions and drawing data-driven conclusions	[Sum]	**NO**	No clear definition; CT is embedded in learning activities like inquiry-based problem-solving, data analysis, and evidence-based reasoning in real-world contexts.

Note: The "Aspects of Critical Thinking Measured" were extracted or inferred from each study. Where possible, phrasing was retained directly from the original source [Orig]; otherwise, entries were summarized [Sum] or interpreted [Inf] based on study content.

APPENDIX C.

Figure 3. PRISMA 2020 Flow Diagram