The role of artificial intelligence in developing critical thinking in higher education students: A systematic review

Guadalupe Martínez Chacón – Universidad Internacional de la Rioja (UNIR) 0000-0002-6301-8828
Salvador Aljazairi López – Universidad Internacional de la Rioja (UNIR) 0000-0002-1564-9860
María Luz Diago-Egaña – Universidad Internacional de la Rioja (UNIR) 0000-0002-0446-3120
Asier Jayo Andrés – Universidad Internacional de la Rioja (UNIR) 0000-0002-9899-8723

Recepción: 15.04.2026 | Aceptado: 19.05.2026

Correspondencia a través de **ORCID**: Guadalupe Martínez Chacón **0000-0002-6301-8828**

Citar: Martínez Chacón, G, Aljazairi López, S, Diago-Egaña, ML, & Jayo Andrés, A (2026). The role of artificial intelligence in developing critical thinking in higher education students: A systematic review. *REIDOCREA, 15*(28), 368-390.
All authors are supported by Universidad Internacional de la Rioja
Disciplinary area: Higher Education

Abstract: The rapid advancement of digital technologies has led to a profound global transformation, reshaping how societies learn, communicate, and produce knowledge. Within this context, artificial intelligence (AI) is increasingly influencing higher education, emerging as both a strategic and practical tool for innovation. This study examines how AI integration affects students' critical thinking. A systematic review of academic literature was conducted, including original articles published in English between 2020 and October 2024. Sources were collected from Web of Science, ProQuest, and Dialnet databases. The initial search identified 341 articles, of which 41 were selected after screening for detailed analysis. Findings reveal that AI integration in higher education generally enhances critical thinking skills, especially when tools are used for text generation, problem-solving, feedback analysis, and information verification. However, the review also emphasizes the importance of cautious implementation, as AI may influence reasoning processes, foster dependency, or introduce inaccuracies if not properly managed. Overall, the evidence suggests that a deliberate and informed use of AI can effectively support the development of critical thinking skills, while also requiring educators to address potential risks associated with its use.)

Keyword: Artificial Intelligence

Introduction

Artificial intelligence (AI) has become one of the most influential technologies of the 21st century, driving profound transformations in science, industry, and society. As a rapidly evolving field, AI offers powerful tools capable of processing information, automating complex tasks, and supporting decision-making at unprecedented scales. Its impact is widely recognized across multiple domains, positioning AI as a cornerstone of contemporary technological development.

The influence of AI extends to a wide range of professional and scientific disciplines. In fields such as engineering, medicine (Hasselgren & Oprea, 2024; Yin et al., 2021), financial sector (Mintarya et al., 2023), customer service (Nicolescu & Tudorache, 2022) , and climate research (Cowls et al., 2021), AI contributes to data analysis, process optimization, and innovative solutions to complex problems. Its multidisciplinary applications highlight the versatility of this technology and its relevance for addressing global challenges.

In recent years, AI has also gained notable prominence in the educational sphere. As educational institutions navigate a rapidly changing technological landscape, AI emerges as a resource capable of rethinking and redesigning teaching and learning processes (Vitanza et al., 2019). Its incorporation into classrooms is accompanied by broader educational aims, such as fostering scientific literacy, strengthening STEM-oriented competencies, and enhancing students' capacity to participate critically in debates about scientific and technological development (Cañal de Leon et al., 2011; Mansour, 2009).

The integration of AI in education aligns with international initiatives such as the European Commission's updated Digital Competence Framework (2022), which incorporates skills and attitudes related to AI and data use. AI offers new possibilities for personalizing learning, supporting teachers, and promoting student autonomy by using tools and platforms supported by Information and Communication Technologies (ICT).

Among the skills potentially enhanced through AI-supported learning, critical thinking (CT) stands out as a central educational objective. CT involves the ability to analyse, interpret, evaluate, and regulate information and reasoning processes (Creamer, 2011; Núñez López et al., 2017). While some studies suggest that AI can stimulate or complement CT development (Chaparro-Banegas et al., 2024), others warn about possible cognitive costs associated with overreliance on AI tools (Moore & Gerlich, 2025). This highlights the importance of understanding how AI is used by students and what implications it has on higher-order cognitive skills.

Research objectives

Given the growing integration of AI in university contexts, it has become essential to understand how its use in academic tasks shapes learning processes and supports the development of key skills. This systematic review therefore examines empirical studies in which higher education students engage with AI tools within instructional settings. The aim is to determine whether these implementations foster the development of CT across different academic disciplines and to deepen our understanding of AI's educational potential as a transformative resource in higher education. To evaluate this, and to explore the factors that may influence these outcomes, we propose the following research questions:

RQ1. How has research on AI and CT in higher education evolved over time, considering both the number of studies published per year and the countries in which they were conducted?

RQ2. What types of learning activities in higher education incorporate students' use of AI tools, and which specific AI tools are employed in these activities?

RQ3. What are the methodological characteristics of the studies and their participants, including sample size, the presence of control and experimental groups, the use of pretest and post-test assessments, study duration, and academic discipline?

RQ4. How is the development of CT assessed in these studies—through quantitative, qualitative, or through mixed-methods approaches?

RQ5. Does the use of AI promote the development of CT, and do the effects on CT vary according to the academic discipline, or the type of learning activities involved?

RQ6. What is the risk of bias of the studies included in the review?

Methods

Protocol and registration: To address the research questions posed in this study, a systematic review of the existing literature on AI and CT was conducted following an explicit, transparent, and structured methodology designed to minimize bias. The review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Liberati et al., 2009), ensuring a rigorous synthesis of the available evidence and the generation of reliable, high-quality conclusions. As this study involved secondary analysis of published data, neither ethical approval nor registration in the International Prospective Register of Systematic Reviews (PROSPERO) was required.

Eligibility criteria: This review was structured beginning with a planning phase in which the research objectives and questions were defined, along with the descriptors to be used and the eligibility criteria established. During this stage, the Population, Intervention, Comparator, Outcome, and Study design (PICOS) framework was applied, in accordance with the PRISMA guidelines, to ensure a systematic and transparent approach to study selection and analysis:

(1) Population: university students of different ages and educational fields are included; (2) Interventions: educational activities in which AI-based tools have been used; (3) Comparator: we included studies involving students who used AI tools, in with CT outcomes were compared either within the same group (pretest–posttest design) or against a control group that did not use AI. (4) Outcome: studies with measurable findings on the development of CT were accepted and (5) Study design: we have included students who have used or applied AI tools in the classroom.

Book chapters, perspective, meeting abstracts, editorials or reviews were excluded. Studies that did not address the use of AI, or that examined the development of CT in teachers or other non-student population, were also excluded.

Information source and search: The studies were extracted from the following three electronic databases: Web of Science, Dialnet, and Proquest. The search was limited to original articles published in English and identified through the following structured search strategy: (TS = artificial intelligence AND critical thinking), covering the period from 2020 to October 2024.

Study selection: Titles and abstracts were screened for inclusion according to the following criteria: studies conducted with students who applied AI tools in the classroom, assessing their CT either through within-group comparisons (pretest–posttest design) or between-group comparisons with students who did not use AI. Theoretical or informative essays, studies that did not involve the use of AI by students, those not conducted with student populations, or those that did not assess CT through objective measures, were excluded. Full-text articles meeting the inclusion criteria were retrieved and evaluated for eligibility. Fig. 1 presents a flow diagram describing the inclusion/exclusion process followed by reasoning.

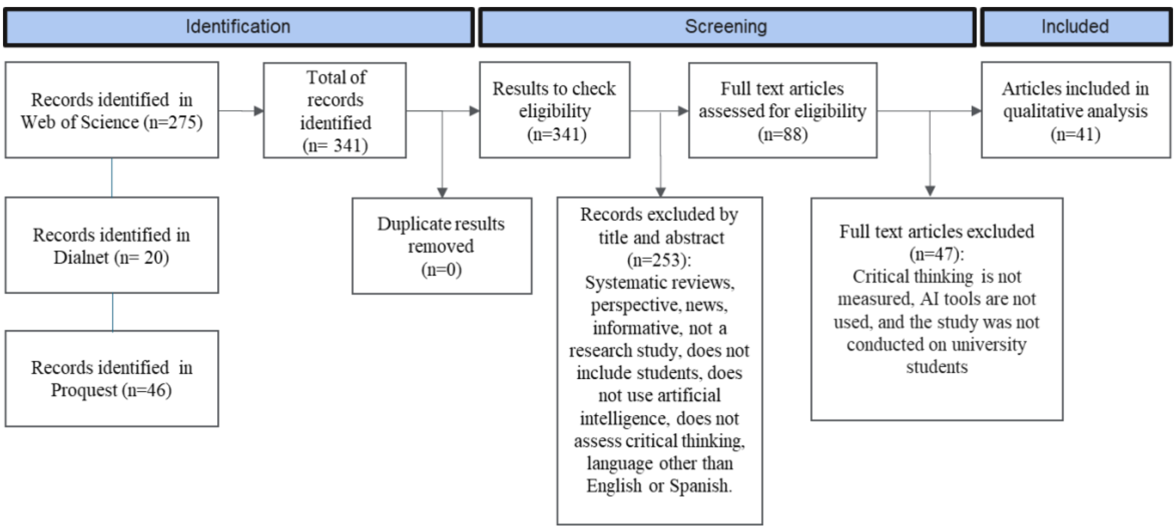
Data collection process: The data collected from the research articles is shown in Table 1. The following data were extracted from the articles: classification of university degrees, degrees, article reference, type of AI used, test or assessment of CT, types of activities using AI, duration of the study and concise conclusion. This information was evaluated, classified, and categorized to elucidate the effect of using AI on the development of CT in university students.

Risk of bias in individual studies/ Reporting bias assessment: To ensure the validity and reliability of the included studies, the methodological risk of bias was independently assessed by the first two authors of this systematic review following the Cochrane Collaboration's tool for assessing risk of bias in randomized trials (Higgins et al., 2011), complemented with additional AI-specific criteria regarding data representativeness, transparency, and reproducibility (Nagendran et al., 2020). The types of bias extracted were selection bias, performance bias, detection bias, attrition bias, reporting bias, and other sources of bias. Specifically, each study examined 9 entries associated with the types of error bias: random sequence generation and allocation concealment (selection bias); knowledge about the use of AI (performance bias) and blinding of assessors (detection bias); completeness of outcome data (attrition bias); selectivity of reported outcomes (reporting bias); and additional potential biases: related to the use of AI and trust in the data obtained (automation bias), data representativeness, transparency, and reproducibility. For each domain, reviewers answered a signalling question and rated the risk of bias as *low*, *high*, or *unclear*. The two reviewers worked independently to determine the adequacy of randomization and allocation concealment, the comparability of baseline characteristics, blinding of participants and outcome assessors, management of incomplete data, and the presence of selective reporting. Any discrepancies in their evaluations were resolved through discussion or consultation with a third reviewer. The final risk of bias classification for each domain was categorized as *low risk*, *high risk*, or *unclear risk*.

Results

A total of 341 potentially relevant original articles were initially identified across the three databases consulted: Web of Science (275), ProQuest (46), and Dialnet (20) (as shown in Figure 1). No duplicate records were found. After screening titles and abstracts based on the predefined eligibility criteria, 253 articles were excluded. Of the remaining 88 full-text articles assessed for eligibility, 47 did not meet the inclusion criteria upon closer examination. This process resulted in a final sample of **41 empirical studies**, reviewed in Fig. 1, which constitute the basis for the analyses presented below.

Fig. 1.
PRISMA flowchart for article selection.



RQ1. How has research on AI and CT in higher education evolved over time, considering both the number of studies published per year and the countries in which they were conducted?

41 original articles published in the period between 2020 and 2024 were selected in this review. Regarding the year of publication: 58.53% of the studies were published in 2024, 36.58% in 2023, 4.87% in 2022 and 2.43% in 2020. The country where the study was carried out is specified in only 36 of the studies (87.80%), form which, 36.58% were conducted in Asia – seven articles in China (He et al., 2023; Li et al., 2024; Shahzad et al., 2024; Xiaolei & Teng, 2024; Yan et al., 2024; Zheng et al., 2024; Zou et al., 2024), four in Taiwan (Chang et al., 2024; P. H. Lin & Chen, 2020; Q Huang et al., 2023; Tu, 2024), and one in Indonesia (Darmawansah et al., 2025), and South Korea (Zhang et

al., 2024)-; 24.39% of the studies were conducted in Europe -two in Germany (Dahlkemper et al., 2023; Henze et al., 2022) and Spain (Murillo-Ligorred et al., 2023; Sánchez-Ruiz et al., 2023), but also in Italy, the United Kingdom, Portugal, France, Romania and the Czech Republic-. 19.51% of the studies were carried out in America, mostly in USA (Favero, 2024; Kim et al., 2024; X. Lin et al., 2024; Mahrous et al., 2023; Wolf & Wolf, 2023), Mexico (Michalon & Camacho-Zuñiga, 2023; Sanabria-Z et al., 2023) and Brazil (Augusto Menten de Barros et al., 2023). Africa was only represented in studies performed in several countries -Argelia, Egypt, Bangladesh, India and Indonesia- (Muthmainnah et al., 2022), and between Nigeria and Bangladesh (Yusuf et al., 2024) . From Oceania (Australia), only one study was selected (Dai et al., 2023).

RQ2. What types of learning activities in higher education incorporate students' use of IA tools, and which specific AI tools are employed in these activities? Among 41 articles, 60.97% (25 articles) involved only one learning activity category, 26.82% (11 articles) involved two different categories, 2.43% (1 article) involved three different categories and 9.75% (4 articles) did not specify the learning activity category employed. The most frequently observed learning activity categories in which AI was used, shown in Table 2, was *text generation*, present in 48.78%% (20 articles) and Writing/performance assessment present in 46.34% (19 articles) and problem solving 14.63% (6 articles). Other activity categories were reported less frequently, such as human–AI interaction for coaching or decision support (Tsopra et al., 2023), AI-supported problem-based learning (Augusto Menten de Barros et al., 2023), paraphrasing or rewriting assistance (Xiaolei & Teng, 2024), automated writing evaluation (Liu et al., 2023), and general generative AI plus diverse applications such as the *PSoC 6 microcontroller technology* (Modran et al., 2024), *Makeblock Neuron Construction Kit* for IoT simulation and 3D-figure manipulation (Henze et al., 2022), *Hereinafter AI* for instructional integration (Murillo-Ligorred et al., 2023), *AiDental* for domain-specific learning support (Mahrous et al., 2023), and an AI platform developed to assess complex thinking (Sanabria-Z et al., 2023). Additionally, augmented learning environments were represented in augmented reality applications (P. H. Lin & Chen, 2020).

RQ3. What are the methodological characteristics of the studies and their participants, including sample size, the presence of control and experimental groups, the use of pretest and post-test assessments, study duration, and academic discipline? A total number of 41 articles were selected, of which 87.80% reported the sample size, ranging from 7 to 1,000 students. Among all analysed studies, 24 conducted both pretests and post-tests; however, only 10 of these 24 included both a control group and an experimental group. Overall, just 24.39% of the studies (n = 10) (Chang et al., 2024; Darmawansah et al., 2025; He et al., 2023; Li et al., 2024; Liu et al., 2023; Mahrous et al., 2023; Huang et al., 2023; Saritepeci & Yildiz Durak, 2024; Yusuf et al., 2024; Zheng et al., 2024) incorporated a control and experimental group in their experimental design.

According to the duration of the study, 58.53% of them indicated the duration of the experiment, which varied between 1 hour [26] and more than 12 months (Xiaolei & Teng, 2024; Zheng et al., 2024), the mode being between 1-3 months (Darmawansah et al., 2025; Li et al., 2024; Mahrous et al., 2023; Michalon & Camacho-Zuñiga, 2023; Muthmainnah et al., 2022; Huang et al., 2023; Saritepeci & Yildiz Durak, 2024; Shahzad et al., 2024).

The academic field more represented in the studies corresponded mainly to Health Sciences (24.39%) (Araujo & Cruz-Correia, 2024; Chang et al., 2024; Favero, 2024; Kämmer et al., 2024; Mahrous et al., 2023;el 9 Tsopra et al., 2023; Wolf & Wolf, 2023;

Zheng et al., 2024), followed by Humanities (21.95%) (philosophy, linguistics, education...) (Dai et al., 2023; Darmawansah et al., 2025; Henze et al., 2022; Hyde et al., 2024; Liu et al., 2023; Modran et al., 2024; Murillo-Ligorred et al., 2023; Saritepeci & Yildiz Durak, 2024; Xiaolei & Teng, 2024; Yan et al., 2024) and Engineering and Architecture (17.07%) (Augusto Menten de Barros et al., 2023; He et al., 2023; Kim et al., 2024; Li et al., 2024; Modran et al., 2024; Q Huang et al., 2023; Sánchez-Ruiz et al., 2023). Only one of the articles correspond to studies in Basic Sciences (2,43%) (Dahlkemper et al., 2023) and another two to Social and Legal Sciences (4.87%) (Essien et al., 2024; Michalon & Camacho-Zuñiga, 2023). The rest of the articles include students from different disciplines in their own studies or not specified (26.82%) (Dindorf et al., 2024; P. H. Lin & Chen, 2020; X. Lin et al., 2024; Muthmainnah et al., 2022; Sanabria-Z et al., 2023; Šedlbauer et al., 2024; Shahzad et al., 2024; Tu, 2024; Yusuf et al., 2024; Zhang et al., 2024; Zou et al., 2024).

RQ4. How is the development of CT assessed in these studies—through quantitative, qualitative, or through mixed-methods approaches? According to the CT measurement tool used in the studies, in 15 out of 41 (36.58%), CT was assessed using a variety of instruments and approaches, mainly of a quantitative nature. These included: a modified Likert scale by Chai et al. (2015); a *Critical Thinking Scale* adapted from the *Motivated Strategies for Learning Questionnaire* (MSLQ) by Pintrich et al. and Pintrich & Others (1991, 1993), in (X. Lin et al., 2024; Liu et al., 2023); Bloom's Taxonomy (Krathwohl, 2002) in Essien et al. (2024) and Murillo-Ligorred et al. (2023); the *California Critical Thinking Skills Test* (CCTS) by Facione & Facione (1992) in (Li et al., 2024; Xiaolei & Teng, 2024; Yusuf et al., 2024); a specific questionnaire by Lin et al. (2019) in (Chang et al., 2024; Liu et al., 2023); the *Clinical Critical Thinking Scale* based on Ennis' (1996) theoretical framework (Zheng et al., 2024); an adapted evaluation rubric (Kim et al., 2024); Kember et al.'s *Reflective Thinking Test* (2000) in Saritepeci & Yildiz Durak (2024) ; a questionnaire adapted in Lin & Chen, 2020; and a questionnaire developed from Miri et al. (2007) in (Muthmainnah et al., 2022). In 18 out of 41 studies analysed (43.90%) evaluation relied on qualitative approaches focused on observation, self-perception, opinions, experiential analysis or fact-checking and reflecting on AI-generated information. In contrast, 8 of the 41 analysed studies (19.51%) did not clearly specify the analytical method used. It was not explicitly indicated whether CT was measured directly, indirectly, or as a dimension within another scale. Nevertheless, various approaches were mentioned: the use of questionnaires to assess participants' attitudes and experiences; integration of CT within the pedagogical approach measurement of academic writing skills; student self-assessments and teacher evaluations of project performance and overall learning within the PBL framework (Augusto Menten de Barros et al., 2023); inclusion of CT as a dimension in the overall assessment of learning and argumentation (He et al., 2023); and the analysis students' relationship with the role of AI tools, rather than as a direct object of evaluation (Favero, 2024).

RQ5. Does the use of AI promote the development of CT, and do the effects on CT vary according to the academic discipline, or the type of learning activities involved? As shown in Table 1, in 35 out of 41 (85.36%) the use of AI enhances the development of CT. Among these 35 articles, 26 involved activities related to text generation and writing, 4 focused on problem-solving, 1 dealt with image generation, 1 involved questioning tasks and 3 of them did not specify the type of AI used. The methods used in these studies to measure the CT varied from quantitative validated measurement scales (fifteen studies), qualitative approaches (twelve studies), and studies in which the analytical methods were not clearly specified (8 studies). Only in 6 out of the 41 reviewed articles (14.63%) (Bonacaro et al., 2024; Dindorf et al., 2024; Hyde et al., 2024; Thompson & Fecher, 2024; Yan et al., 2024; Zhang et al., 2024) the main conclusion indicates that the application of AI tends to limit CT and to reduce its potential for improvement. Furthermore, in these studies it is noted that AI use may create a certain degree of dependency among students and requires an adaptation period for effective use. It is interesting to note that all these studies shared a CT evaluation methodology that relied on qualitative approaches focused on observation, self-perception, opinions, experiential analysis or fact-checking and reflecting on AI-generated information, and that none of them used a quantitative validated measurement scale. Considering the academic field of the studies, quantitative CT measurement probed an increase in a variety of subjects ranging from Humanities (5 studies), mixed-disciplines (4), engineering (3) and social sciences (3). The studies that measured an increase in CT with qualitative tools likewise covered Science (1), Humanities (1), Health Sciences (3), Social Sciences (1), Engineering (2), and Multiple Disciplines (4).

RQ5. *What is the risk of bias of the studies* included in the review? To assess the risk of bias, the Cochrane Collaboration's tool for assessing risk of bias in randomized trials (Higgins et al., 2011) was applied, as presented in table 3. Each domain was examined, and the level of risk was classified as unclear, low or high. Analysed and the risk value was established as unclear, low or high. In most domains, the risk of bias was judged to be unclear or high, due to the limited availability of relevant methodological information in the included studies. However, in the domain of automation bias, the risk of bias was predominantly assessed as low, a finding of particular relevance for this systematic review, as it acknowledges that AI-generated data may contain inaccuracies, and this fact could potentially stimulate the development of CT among students.

Selection bias: According to random sequence, in 12 of the 41 articles (29.6%), group assignment was randomised; in 14 of the 41 articles (34.14%), the assignment of students to each experimental group was not randomised; in 12 of the 42 articles (29.26%), it was not specified; and in the remaining 3 studies, it was not applicable. In relation to allocation concealment in 12 of the 41 articles (29.26%), it is stated that all participants were informed about the purpose of the experiment and participated voluntarily, and the study group was known. However, in 20 of the 41 articles (48.78%), it is not explicitly stated whether the students knew which specific group they belonged to. Furthermore, in 2 of the studies did not disclose the group to which the students belonged; in 7 cases, this is not applicable (16.6%).

Performance bias: Regarding the assessments, in 6 articles (14.63%) the evaluators were unaware of the group to which the students belonged; in 24 articles (58.53%) it was not specified; in 3 studies (7,31%) it was not applicable; in 8 articles (19.51%) the evaluators were aware of the group characteristics.

Detection bias: Furthermore, when analyzing the results, only 7 studies (17.07%) were blinded; in 22 publications (53.65%), blinding procedure was not described; in 9 articles (21.95%), the evaluators were aware of the conditions; and in 3 articles (7.31%), it was not applied.

Attrition and reporting bias: In 19 of the studies (46.34%), dropout rates are recorded, in 5 of the studies (11.90%) the description of this rate is unclear, and in 17 of the studies (40.47%) no dropouts are reported. Of the 20 cases with dropout rates recorded, in 4 articles (9.52%) it is unclear whether unexpected data or only success rates are being recorded.

Automation bias: In 31 of the articles (75.60%) acknowledge that AI can sometimes provide incorrect feedback and sometimes provide additional feedback on the content. In 7 of the articles (17.07%), authors do not mention this distrust of the data or its identification is unclear. In 3 studies (7.31%), they do not address whether AI-generated data is reliable even if it is incorrect, nor how potential inaccuracies are handled.

Representative data: In 20 of the 41 articles (8.78%), the number of students may not be representative, due to a low sample size, low participation rate, students from very specific disciplines, or demographic characteristics. In 21 (51.21%) of the articles, the representativeness of the data used is not stated or is not entirely clear. Regarding transparency and reproducibility, 6 studies (14.63%) provide details on their methodology, including the development and validation of the proposed framework, data collection procedures, and methods; 14 studies (34.14%) provide unclear details as not all aspects are specified; and 2 articles (4,87%) provide low levels of reproducibility and transparency, as these aspects are not fully specified or not mentioned, and the methodology is not clearly described. In the remaining studies, we found 11 (Low/High), 5 (Low/Unclear), and 3 (Unclear/High) regarding reproducibility and transparency.

Discussion

The results of this systematic review show a clear increase in the use of AI tools in education from 2020 until November 2024, with most publications (95.11%) being published between 2023 and 2024. This growth reflects the rapid adoption of generative technologies and AI-based learning support systems (Pang & Wei, 2025). This finding is consistent with other studies showing that generative AI has become a popular educational resource to support teaching and learning in higher education (Batista et al., 2024). Furthermore, the integration of AI in education has the potential to revolutionize how students learn and interact with information (Labadze et al., 2023). Following this global trend, models such as ChatGPT (AL-Smadi, 2023) have gained particular prominence, being identified as the most frequently used tool in our study (60.97%).

The geographic distribution of the studies on IA with educational purposes reveals a higher concentration in Asia (36,58%), particularly in China and Taiwan, followed by Europe (24.36%) and the Americas (mainly the United States and Brazil). These data are consistent with recent bibliometric studies and reviews showing a rapid expansion of research in educational AI, with outstanding contributions from China and the US and a notable increase in publications on generative AI (Ng & Ho, 2025). This pattern suggests that research activity is concentrated in regions with strong investment in technological innovation (Britto et al., 2015) and educational research (Huang et al., 2014; Knight, 2022). However, the limited number of studies from Oceania, Latin America and Africa highlights the need for more diverse and context-sensitive research, since the effectiveness and acceptance of AI may vary depending on cultural, economic, and pedagogical factors. However, AI could be a big opportunity to develop innovation in educational research because many ai technologies are available to everyone in countries as for example, in Latin America (Pedro et al., 2019).

In this context, chatbot technology stands out as one of the most widely adopted and researched AI applications in education, with a significant impact on overall learning outcomes. Specifically, chatbots have demonstrated notable improvements in learning achievement, explicit reasoning, and knowledge retention. Their integration offers multiple benefits, including immediate assistance, quick access to information, and enhanced educational experiences. However, research findings remain mixed regarding their influence on CT, engagement, and motivation (Deng & Yu, 2023) found that chatbots positively affect several learning-related aspects but do not significantly improve student motivation. In contrast, other studies (Okonkwo & Ade-Ibijola, 2021; Wollny et al., 2021) suggest that chatbot use can increase motivation among learners, highlighting the complexity and contextual nature of AI's pedagogical impact.

Methodologically, only 24,39% of the studies included both control and experimental groups. Together with this, the results obtained in error bias showed that 70,78% of the studies did not specify whether the experimental groups were randomised, and less than 60% reported the duration of their interventions. These methodological limitations make it difficult to draw strong conclusions about the real effectiveness of AI tools in learning outcomes. The wide variation in study designs and intervention durations—from one-hour sessions to programs lasting over a year—reflects both the flexibility and the lack of standardization in the educational use of AI (Zawacki-Richter et al., 2019) In this sense, Reed et al. (2005) highlighted that systematic reviews require comparable methodologies, the presence of control groups, and similar types of interventions to achieve more equivalent and reliable results. In the present review, the diversity of designs and experimental conditions limits the ability to generalize findings and to compare the true impact of AI tools across educational contexts. Moreover, only six studies provided detailed methodological information, indicating transparency and favouring reproducibility. The remaining articles showed low levels risk of bias of reproducibility and transparency, as these aspects were not fully specified and their methodologies were not clearly described. Future research should aim for stronger experimental designs and clearer reporting to better understand how AI impacts student performance and cognitive skills.

The study of science and other fields of knowledge must be carried out with an understanding of the context in which they emerged and the processes that shaped them. This is particularly relevant considering that the history of educational systems is deeply connected to the development of industrial society, and that many current learning practices continue to respond to the needs of an ever-evolving technological and social environment. Within this broader context, recent research shows how different academic fields are integrating AI in diverse yet converging ways. Most studies have been conducted in the Health Sciences (26.82%), and 81% of them suggest that AI is perceived as a valuable tool in areas where decision-making and CT are essential (Reed et al., 2005). Although less common, studies in the Humanities and Engineering also highlight AI's broad potential to promote independent learning, enhance academic writing, and support complex problem-solving (Pedro et al., 2019). Taken together, this evidence reinforces the need to integrate AI into educational processes, acknowledging both its disciplinary impact and the historical and social evolution of learning models.

The variety of tools identified—from *ChatGPT* and *Quillbot* to more specific systems like *AI-PBL* or *AI-CDSS*—shows the diversity of pedagogical goals and approaches in AI integration (Kayal, 2024). On one hand, this wide variety complicates data systematization, but on the other hand, it opens a wide range of learning opportunities due to the wide variety of resources AI offers, and it also opens a diversity of needs for this new learning (Kayal, 2024).

The role of AI on the development of CT was studied across the present review through the analysis of 41 studies. The CT provides keys to understanding the progression of the basic processes to learn in higher levels of education (Martínez et al., 2025). Across the analysed studies, CT assessment displayed substantial methodological diversity. Fifteen studies employed a myriad of instruments, from well-established and validated questionnaires to purpose-built tools for each study. This methodological diversity illustrates a strong trend toward operationalizing CT as a measurable construct (Carter et al., 2015). However, the use of distinct instruments across studies raises concerns about comparability and construct validity, since different tools capture different CT dimensions—skills, dispositions, or behaviours (Liang, 2020; Liao et al., 2025). In addition, twelve studies (29.26%) assessed CT using qualitative approaches, focusing on observation, perception, and experiential analysis. Although these approaches capture richer contextual insights, they often lack replicability or standardized interpretation (Sternberg, 2024). Moreover, eight studies failed to specify their analytical methods clearly, leaving it uncertain whether CT was measured directly, indirectly, or embedded within broader learning scales. Such methodological ambiguity limits confidence in the resulting conclusions (Zhai et al., 2024).

One interesting finding of this review is that in 85.36% of the analysed studies, integrating AI into educational processes is associated with a positive trend in the development of CT (Liang, 2020). Regarding one of the additional potential biases related to the use of AI and trust in the data obtained (automation bias), in 31 of the articles (approximately 75%), students identified errors that the machine missed or provided more detailed feedback. This indicates an awareness that AI-generated data can be inaccurate and fosters CT among students. It also shows that AI-generated data was not always reliable and that, when errors occurred, students developed strategies to verify the accuracy of the information, further promoting the development of CT, as reported in 24 of the 31 cases. Only 6 of the 41 studies (14.63%) concluded that AI can limit CT, but interestingly, none of these studies used any quantitative, previously validated CT measurement scale, something that has to be considered in further studies is whether or not the results observed are dependent on the measurement method used, related to the detection bias.

The findings that these studies share, suggest that AI integration can support CT when implemented deliberately, generating dynamic and interactive learning experiences that encourage active participation (Baidoo-Anu & Owusu Ansah, 2023). Nevertheless, the extent of their effectiveness is contingent upon the existence of structured frameworks that inform instructional design and facilitate the assessment of genuine skill development (Martínez et al., 2025). This AI integration in educational processes may be also risking over-dependence or reduced cognitive engagement if misused (Delgado et al., 2015) and also, requires an adaptation period for effective use. Therefore, the positive or negative impact of AI on CT appears to depend less on the technology itself and more on how it is pedagogically integrated.

Regarding disciplinary context, studies analysing CT with quantitative and qualitative tools showed an improvement of CT skills along different disciplines. These distributions suggest that the potential for CT development via educational interventions is broad and not confined to any single field. CT development through AI studies across many disciplines also suggest that AI-supported active learning benefits diverse academic areas (Torrisi-Steele, 2026).

Overall, the findings of this review indicate a rapidly expanding but still developing field. AI is being used for multiple purposes—automated assessment, text generation, writing support, and problem-solving—but empirical evidence of its educational impact remains limited and methodologically inconsistent. As previously discussed, further developments of this tool in education can potentially foster CT improvement in students, as it also raises questions such as: are teachers ready to transition to an AI-integrated educational process? (Luckin et al., 2022) How must we modify the subject`s curriculum to be ready for IA education? (Sai Madhavi Palle, 2025) How can we prepare our student for an AI driven education? (George, 2023). Future research should prioritize rigorous comparative designs, include longitudinal measures, and address ethical and pedagogical aspects associated with the use of AI in learning environments.

In summary, the present review offers encouraging evidence that AI-supported educational interventions can enhance CT if measurement is rigorous, pedagogy is thoughtful, and students are actively engaged in higher-order cognitive tasks. The challenge for educators is to harness AI in ways that *stimulate* rather than *substitute* the thinking process. However, the heterogeneity of CT measures and evaluation procedures constrains cross-study comparability. Future research should prioritize standardization of CT assessment tools and longitudinal designs to track sustained effects of AI-supported in education. Additionally, the rapid evolution of generative AI warrants continuous examination of how new tools reshape cognitive processes in education.

Conclusions

Evidence suggests that AI-supported active learning fosters the development of CT and other higher-order skills across different disciplines, indicating that its benefits are not limited to a single academic field. However, the impact of AI depends less on the technology itself and more on how it is pedagogically implemented.

The growing use of AI in education reflects the rapid adoption of generative technologies and AI-based learning systems that are transforming teaching and learning. While this expansion demonstrates AI's potential to enhance educational practices, the diversity of research designs and contexts still limits the comparability and generalization of findings. Future studies should therefore employ stronger experimental designs and clearer reporting to better understanding of AI effects on student learning and cognitive development.

Ultimately, as AI continues to evolve, it will play an increasingly influential role in shaping educational reform, driving new ideas, methods, and approaches that redefine how learners think, create, and solve problems.

References

AL-Smadi, M. (2023). ChatGPT and Beyond: The Generative AI Revolution in Education. http://arxiv.org/abs/2311.15198

Araujo, S. M., & Cruz-Correia, R. (2024). Incorporating ChatGPT in Medical Informatics Education: Mixed Methods Study on Student Perceptions and Experiential Integration Proposals. JMIR Medical Education, 10. https://doi.org/10.2196/51151

Augusto Menten de Barros, V., Mohallem Paiva Senior Member, H., & Takashi Hayashi, V. (2023). Using PBL and Agile to Teach Artificial Intelligence to Undergraduate Computing Students; Using PBL and Agile to Teach Artificial Intelligence to Undergraduate Computing Students. IEEE Access, PP. https://doi.org/10.1109/ACCESS.2017.DOI

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. In Journal of AI (Vol. 52, Issue 7).

Başol, G., & Evin-Gencel. (2013). Yansıtıcı Düşünme Düzeyini Belirleme Ölçeği | TOAD. Educational Sciences: Theory & Practice (ESTP). https://toad.halileksi.net/olcek/yansitici-dusunme-duzeyini-belirleme-olcegi/

Batista, J., Mesquita, A., & Carnaz, G. (2024). Generative AI and Higher Education: Trends, Challenges, and Future Directions from a Systematic Literature Review. In Information (Switzerland) (Vol. 15, Issue 11). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/info15110676

Bonacaro, A., Rubbi, I., Artioli, G., Monaco, F., Sarli, L., & Guasconi, M. (2024). AI and Big Data: Current and Future Nursing Practitioners' Views on Future of Healthcare Education Provision. Studies in Health Technology and Informatics, 315, 200–204. https://doi.org/10.3233/SHTI240134

Britto, G., Pereira, U., Santos, D., Kruss, G., & Albuquerque, E. (2015). Global innovation networks and university-firm interactions: an exploratory survey analysis. Revista Brasileira de Inovação, 14(1), 163–192. https://doi.org/10.20396/RBI.V14I1.8649093

Cañal de Leon, P. ;, Barberá, Ó., Brusi, D., Fernández Manzanal, R., Gil, C., Gil, D., Giordan, A., Aránzazu Hueto, M., Membiela, P., Pedrinaci, E., Rivero, A., Sendra, C., Vilches, A., María Wamba Pedro Cañal, A., Gil IES Berenguer Dalmau Catarroja, C., Aránzazu Hueto IES Avempace Zaragoza Pedro Membiela, M., Pedrinaci IES El Majuelo Gines Ana Rivero, E., & María Wamba, A. (2011). Los enfoques integrados de Ciencia-Tecnología-Sociedad en la enseñanza secundaria. Biología y Geología: Complementos de Formación Disciplinar, 2011, ISBN 978-84-9980-049-3, Págs. 123-142, 123–142. https://dialnet.unirioja.es/servlet/articulo?codigo=3659185

Carter, A. G., Creedy, D. K., & Sidebotham, M. (2015). Evaluation of tools used to measure critical thinking development in nursing and midwifery undergraduate students: A systematic review. In Nurse Education Today (Vol. 35, Issue 7, pp. 864–874). Churchill Livingstone. https://doi.org/10.1016/j.nedt.2015.02.023

Chai, C. S., Deng, F., Tsai, P. S., Koh, J. H. L., & Tsai, C. C. (2015). Assessing multidimensional students' perceptions of twenty-first-century learning practices. Asia Pacific Education Review, 16(3), 389–398. https://doi.org/10.1007/s12564-015-9379-4

Chang, C. Y., Yang, C. L., Jen, H. J., Ogata, H., & Hwang, G. H. (2024). Facilitating nursing and health education by incorporating ChatGPT into learning designs. Educational Technology and Society, 27(1), 215–230. https://doi.org/10.30191/ETS.202401_27(1).TP02

Chaparro-Banegas, N., Mas-Tur, A., & Roig-Tierno, N. (2024). Challenging critical thinking in education: new paradigms of artificial intelligence. Cogent Education, 11(1). https://doi.org/10.1080/2331186X.2024.2437899;WGROUP:STRING:PUBLICATION

Chen, C.-C. (2000). An experimental study about implementing the teaching strategies of critical thinking in social class of 5th-grade elementary school. (Unpublished Master's Thesis). National Chengchi University. . https://thesis.lib.nccu.edu.tw/cgi-bin/gs32/gsweb.cgi/ccd=NoqOuo/record?r1=1&h1=1

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. AI & SOCIETY 2021 38:1, 38(1), 283–307. https://doi.org/10.1007/S00146-021-01294-X

Creamer, M. (2011). ¿Qué es y por qué pensamiento crítico? In Curso de didáctica del pensamiento crítico (pp. 11–22). Ecuador, Ministerio de Educación.

Dahlkemper, M. N., Lahme, S. Z., & Klein, P. (2023). How do physics students evaluate artificial intelligence responses on comprehension questions A study on the perceived scientific accuracy and linguistic quality of ChatGPT. Physical Review Physics Education Research, 19(1). https://doi.org/10.1103/PhysRevPhysEducRes.19.010142

Dai, Y., Lai, S., Lim, C. P., & Liu, A. (2023). ChatGPT and its impact on research supervision: Insights from Australian postgraduate research students. In Australasian Journal of Educational Technology (Vol. 2023, Issue 4).

Darmawansah, D., Rachman, D., Febiyani, F., & Hwang, G. J. (2025). ChatGPT-supported collaborative argumentation: Integrating collaboration script and argument mapping to enhance EFL students' argumentation skills. Education and Information Technologies, 30(3), 3803–3827. https://doi.org/10.1007/s10639-024-12986-4

Delgado, A. J., Wardlow, L., Mcknight, K., & Malley, K. (2015). Educational technology: A review of the integration, resources, and effectiveness of technology in K-12 classrooms. In Journal of Information Technology Educa-tion: Research (Vol. 14). http://www.jite.org/documents/Vol14/JITEv14ResearchP397-416Delgado1829.pdf

Deng, X., & Yu, Z. (2023). A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education. Sustainability (Switzerland), 15(4), 2940. https://doi.org/10.3390/SU15042940/S1

Dindorf, C., Weisenburger, F., Bartaguiz, E., Dully, J., Klappenberger, L., Lang, V., Zimmermann, L., Fröhlich, M., & Seibert, J. N. (2024). Exploring Decision-Making Competence in Sugar-Substitute Choices: A Cross-Disciplinary Investigation among Chemistry and Sports and Health Students. Education Sciences, 14(5). https://doi.org/10.3390/educsci14050531

Ennis, R. Hugh. (1996). Critical thinking. Prentice Hall.

Essien, A., Bukoye, O. T., O'Dea, X., & Kremantzis, M. (2024). The influence of AI text generators on critical thinking skills in UK business schools. Studies in Higher Education, 49(5), 865–882. https://doi.org/10.1080/03075079.2024.2316881

Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: the development of the California Critical Thinking Disposition Inventory. The Journal of Nursing Education, 33(8), 345–350. https://doi.org/10.3928/0148-4834-19941001-05

Facione, P. A., & Facione, N. C. (1992). The California Critical Thinking Disposition Inventory. California Academic Press.

Favero, T. G. (2024). Using artificial intelligence platforms to support student learning in physiology. Advances in Physiology Education, 48(2), 193–199. https://doi.org/10.1152/advan.00213.2023

George, A. S. (2023). Preparing Students for an AI-Driven World: Rethinking Curriculum and Pedagogy in the Age of Artificial Intelligence. Partners Universal Innovative Research Publication (PUIRP) Volume: 01 Issue: 02 | November-December 2023 | Www.Puirp.Com, 01(02), 112–136. https://doi.org/10.5281/ZENODO.10245675

Hasselgren, C., & Oprea, T. I. (2024). Artificial Intelligence for Drug Discovery: Are We There Yet? Annual Review of Pharmacology and Toxicology, 64(Volume 64, 2024), 527–550. https://doi.org/10.1146/ANNUREV-PHARMTOX-040323-040828/CITE/REFWORKS

He, J., Ma, T., & Zhang, Y. (2023). Design of blended Learning Mode and Practice Community using Intelligent Cloud Teaching. Education and Information Technologies, 28(8), 10593–10615. https://doi.org/10.1007/s10639-023-11606-x

Henze, J., Schatz, C., Malik, S., & Bresges, A. (2022). How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training? Frontiers in Education, 7. https://doi.org/10.3389/feduc.2022.872637

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ (Online), 343(7829). https://doi.org/10.1136/bmj.d5928

Huang, F., Teichler, U., & Galaz-Fontes, J. F. (2014). Regionalisation of Higher Education and the Academic Profession in Asia, Europe and North America. The Internationalization of the Academy: Changes, Realities and Prospects, 145–181. https://doi.org/10.1007/978-94-007-7278-6_8

Hyde, S. J., Busby, A., & Bonner, R. L. (2024). Tools or Fools: Are We Educating Managers or Creating Tool-Dependent Robots? Journal of Management Education, 48(4), 708–734. https://doi.org/10.1177/10525629241230357

Kämmer, J. E., Hautz, W. E., Krummrey, G., Sauter, T. C., Penders, D., Birrenbach, T., & Bienefeld, N. (2024). Effects of interacting with a large language model compared with a human coach on the clinical diagnostic process and outcomes among fourth-year medical students: Study protocol for a prospective, randomised experiment using patient vignettes. BMJ Open, 14(7). https://doi.org/10.1136/bmjopen-2024-087469

Kayal, A. (2024). Transformative Pedagogy: A Comprehensive Framework for AI Integration in Education. 247–270. https://doi.org/10.1007/978-3-031-72410-7_14

Kember, D., Leung, D. Y. P., Jones, A., Loke, A. Y., McKay, J., Sinclair, K., Tse, H., Webb, C., Wong, F. K. Y., Wong, M., & Yeung, E. (2000). Development of a questionnaire to measure the level of reflective thinking. Assessment and Evaluation in Higher Education, 25(4), 381–395. https://doi.org/10.1080/713611442;WGROUP:STRING:PUBLICATION

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, 7(2), 259–289. https://doi.org/10.46328/ijte.691

Knight, J. (2022). The Changing World of International Higher Education, Research and Innovation. Knowledge Diplomacy in International Relations and Higher Education, 39–50. https://doi.org/10.1007/978-3-031-14977-1_4

Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). Computers in Human Behavior, 72, 558–569. https://doi.org/10.1016/J.CHB.2017.01.005

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview.

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. In International Journal of Educational Technology in Higher Education (Vol. 20, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s41239-023-00426-1

Li, T., Ji, Y., & Zhan, Z. (2024). Expert or machine? Comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course. Asia Pacific Journal of Education, 44(1), 45–60. https://doi.org/10.1080/02188791.2024.2305163

Liang, W. (2020). Development Trend and Thinking of Artificial Intelligence in Education. 2020 International Wireless Communications and Mobile Computing, IWCMC 2020, 886–890. https://doi.org/10.1109/IWCMC48107.2020.9148078

Liao, W., Liu, W., & Yuan, R. (2025). How is teachers' critical thinking defined, approached, measured, and evaluated in empirical studies? A methodological review. Thinking Skills and Creativity, 57. https://doi.org/10.1016/j.tsc.2025.101795

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. In PLoS Medicine (Vol. 6, Issue 7). https://doi.org/10.1371/journal.pmed.1000100

Lin, P. H., & Chen, S. Y. (2020). Design and Evaluation of a Deep Learning Recommendation Based Augmented Reality System for Teaching Programming and Computational Thinking. IEEE Access, 8, 45689–45699. https://doi.org/10.1109/ACCESS.2020.2977679

Lin, X., Luterbach, K., Gregory, K. H., & Sconyers, S. E. (2024). A Case Study Investigating the Utilization of ChatGPT in Online Discussions. Online Learning Journal, 28(2). https://doi.org/10.24059/olj.v28i2.4407

Lin, Y. N., Hsia, L. H., Sung, M. Y., & Hwang, G. H. (2019). Effects of integrating mobile technology-assisted peer assessment into flipped learning on students' dance skills and self-efficacy. Interactive Learning Environments, 27(8), 995–1010. https://doi.org/10.1080/10494820.2018.1461115;ISSUE:ISSUE:DOI

Liu, C. C., Liu, S. J., Hwang, G. J., Tu, Y. F., Wang, Y., & Wang, N. (2023). Engaging EFL students' critical thinking tendency and in-depth reflection in technology-based writing contexts: A peer assessment-incorporated automatic evaluation approach. Education and Information Technologies 2023 28:10, 28(10), 13027–13052. https://doi.org/10.1007/S10639-023-11697-6

Luckin, R., Cukurova, M., Kent, C., & du Boulay, B. (2022). Empowering educators to be AI-ready. Computers and Education: Artificial Intelligence, 3. https://doi.org/10.1016/j.caeai.2022.100076

Mahrous, A., Botsko, D. L., Elgreatly, A., Tsujimoto, A., Qian, F., & Schneider, G. B. (2023). The use of artificial intelligence and game-based learning in removable partial denture design: A comparative study. Journal of Dental Education, 87(8), 1188–1199. https://doi.org/10.1002/jdd.13225

Mansour, N. (2009). Science-Technology-Society (STS). Bulletin of Science, Technology & Society, 29(4), 287–297. https://doi.org/10.1177/0270467609336307

Martínez, C. M., Roger-Monzo, V., & Castelló-Sirvent, F. (2025). Generative AI and critical thinking in online higher education: challenges and opportunities. RIED-Revista Iberoamericana de Educacion a Distancia, 28(2), 233–273. https://doi.org/10.5944/ried.28.2.43556

Michalon, B., & Camacho-Zuñiga, C. (2023). ChatGPT, a brand-new tool to strengthen timeless competencies. Frontiers in Education, 8. https://doi.org/10.3389/feduc.2023.1251163

Mintarya, L. N., Halim, J. N. M., Angie, C., Achmad, S., & Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. Procedia Computer Science, 216, 96–102. https://doi.org/10.1016/J.PROCS.2022.12.115

Miri, B., David, B. C., & Uri, Z. (2007). Purposely Teaching for the Promotion of Higher-order Thinking Skills: A Case of Critical Thinking. Research in Science Education 2007 37:4, 37(4), 353–369. https://doi.org/10.1007/S11165-006-9029-2

Modran, H. A., Ursuțiu, D., & Samoilă, C. (2024). Using the Theoretical-Experiential Binomial for Educating AI-Literate Students. Sustainability (Switzerland), 16(10). https://doi.org/10.3390/su16104068

Moore, L. J., & Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies 2025, Vol. 15, Page 6, 15(1), 6. https://doi.org/10.3390/SOC15010006

Murillo-Ligorred, V., Ramos-Vallecillo, N., Covaleda, I., & Fayos, L. (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). https://doi.org/10.3390/educsci13111073

Muthmainnah, Ibna Seraj, P. M., & Oteir, I. (2022). Playing with AI to Investigate Human-Computer Interaction Technology and Improving Critical Thinking Skills to Pursue 21stCentury Age. Education Research International, 2022. https://doi.org/10.1155/2022/6468995

Nagendran, M., Chen, Y., Lovejoy, C. A., Gordon, A. C., Komorowski, M., Harvey, H., Topol, E. J., Ioannidis, J. P. A., Collins, G. S., & Maruthappu, M. (2020). Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies in medical imaging. The BMJ, 368. https://doi.org/10.1136/bmj.m689

Ng, S. L., & Ho, C. C. (2025). Generative AI in Education: Mapping the Research Landscape Through Bibliometric Analysis. Information (Switzerland), 16(8). https://doi.org/10.3390/info16080657

Nicolescu, L., & Tudorache, M. T. (2022). Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review. Electronics 2022, Vol. 11, Page 1579, 11(10), 1579. https://doi.org/10.3390/ELECTRONICS11101579

Núñez López, S., Ávila Palet, J. E., & Olivares Olivares, S. L. (2017). El desarrollo del pensamiento crítico en estudiantes universitarios por medio del Aprendizaje Basado en Problemas. Revista Iberoamericana de Educación Superior, ISSN-e 2007-2872, Vol. 8, Nº. 23, 2017, Págs. 84-103, 8(23), 84–103. https://dialnet.unirioja.es/servlet/articulo?codigo=8866347&info=resumen&idioma=ENG

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. Computers and Education: Artificial Intelligence, 2, 100033. https://doi.org/10.1016/J.CAEAI.2021.100033

Pang, W., & Wei, Z. (2025). Shaping the Future of Higher Education: A Technology Usage Study on Generative AI Innovations. In Information (Switzerland) (Vol. 16, Issue 2). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/info16020095

Payan-Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., Pnevmatikos, D., Payan-Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., & Pnevmatikos, D. (2022). Development and Validation of a Critical Thinking Assessment-Scale Short Form. Education Sciences 2022, Vol. 12, 12(12). https://doi.org/10.3390/EDUCSCI12120938

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development. Roscongress Building Trust.

Pintrich, P. R., & Others, A. (1991). A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ).

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). Educational and Psychological Measurement, 53(3), 801–813. https://doi.org/10.1177/0013164493053003024

Q Huang, A. Y., Wei Chang, J., M Yang, A. C., Ogata, H., Ting Li, S., Xuan Yen, R., & H Yang, S. J. (2023). Personalized Intervention based on the Early Prediction of At-risk Students to Improve Their Learning Performance. Technology & Society, 26(4), 69–89. https://doi.org/10.2307/48747521

Reed, D., Price, E. G., Windish, D. M., Wright, S. M., Gozu, A., Hsu, E. B., Beach, M. C., Kern, D., & Bass, E. B. (2005). Challenges in Systematic Reviews of Educational Intervention Studies. Https://Doi.Org/10.7326/0003-4819-142-12_Part_2-200506211-00008, 142(12 II), 1080–1089. https://doi.org/10.7326/0003-4819-142-12_PART_2-200506211-00008

Sai Madhavi Palle. (2025). Teachers readiness to integrate AI in Education (AIEd) AI literacy of K-12 teachers in India and its influence on their readiness to integrate AIEd.

Sanabria-Z, J., Castillo-Martínez, I. M., González-Pérez, L. I., & Ramírez-Montoya, M. S. (2023). Complex thinking through a Transition Design-guided Ideathon: testing an AI platform on the topic of sharing economy. Frontiers in Education, 8. https://doi.org/10.3389/feduc.2023.1186731

Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. Applied Sciences (Switzerland), 13(10). https://doi.org/10.3390/app13106039

Saritepeci, M., & Yildiz Durak, H. (2024). Effectiveness of artificial intelligence integration in design-based learning on design thinking mindset, creative and reflective thinking skills: An experimental study. Education and Information Technologies, 29(18), 25175–25209. https://doi.org/10.1007/s10639-024-12829-2

Šedlbauer, J., Činčera, J., Slavík, M., & Hartlová, A. (2024). Students' reflections on their experience with ChatGPT. Journal of Computer Assisted Learning, 40(4), 1526–1534. https://doi.org/10.1111/jcal.12967

Shahzad, M. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: the role of trust as a cornerstone. International Journal of Educational Technology in Higher Education, 21(1). https://doi.org/10.1186/s41239-024-00478-x

Simms, R. C. (2024). Work With ChatGPT, Not Against: 3 Teaching Strategies That Harness the Power of Artificial Intelligence. Nurse Educator, 49(3), 158–161. https://doi.org/10.1097/NNE.0000000000001634

Sternberg, R. J. (2024). Do Not Worry That Generative AI May Compromise Human Creativity or Intelligence in the Future: It Already Has. Journal of Intelligence 2024, Vol. 12, Page 69, 12(7), 69. https://doi.org/10.3390/JINTELLIGENCE12070069

Thompson, S. H., & Fecher, P. (2024). Using Adult Learning Theory to Guide Instruction in Training Physical Therapy Students on the Use of AI in Didactic Education. Medical Science Educator, 34(5), 971–973. https://doi.org/10.1007/s40670-024-02066-0

Torrisi-Steele, G. (2026). AI in Higher Education: Opportunities and Challenges. . In In V. Wang (Ed.), Foundations and Frameworks for AI in Education (pp. 171–198). IGI Global Scientific Publishing.

Tsopra, R., Peiffer-Smadja, N., Charlier, C., Campeotto, F., Lemogne, C., Ruszniewski, P., Vivien, B., & Burgun, A. (2023). Putting undergraduate medical students in AI-CDSS designers' shoes: An innovative teaching method to develop digital health critical thinking. International Journal of Medical Informatics, 171. https://doi.org/10.1016/j.ijmedinf.2022.104980

Tu, Y. F. (2024). Roles and functionalities of ChatGPT for students with different growth mindsets: Findings of drawing analysis. Educational Technology and Society, 27(1), 198–214. https://doi.org/10.30191/ETS.202401_27(1).TP01

Vitanza, A., Rossetti, P., Mondada, F., & Trianni, V. (2019). Robot swarms as an educational tool: The Thymio's way. International Journal of Advanced Robotic Systems, 16(1). https://doi.org/10.1177/1729881418825186;REQUESTEDJOURNAL:JOURNAL:ARXA;WEBSITE:WEBSITE:SAGE;WGROUP:STRING:PUBLICATION

Wilson, K. L., Lizzio, A., & Ramsden, P. (1997). The Development, Validation and Application of the Course Experience Questionnaire. Studies in Higher Education, 22(1), 33–53. https://doi.org/10.1080/03075079712331381121;PAGE:STRING:ARTICLE/CHAPTER

Wolf, R. R., & Wolf, A. B. (2023). Using AI to Evaluate a Competency-Based Online Writing Course in Nursing. Online Learning Journal, 27(3), 41–69. https://doi.org/10.24059/olj.v27i3.3974

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are We There Yet? - A Systematic Literature Review on Chatbots in Education. Frontiers in Artificial Intelligence, 4, 654924. https://doi.org/10.3389/FRAI.2021.654924/ENDNOTE

Xiaolei, S., & Teng, M. F. (2024). Three-wave cross-lagged model on the correlations between critical thinking skills, self-directed learning competency and AI-assisted writing. Thinking Skills and Creativity, 52. https://doi.org/10.1016/j.tsc.2024.101524

Yan, Y., Sun, W., & Zhao, X. (2024). Metaphorical conceptualizations of generative artificial intelligence use by Chinese university EFL learners. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1430494

Yin, J., Ngiam, K. Y., & Teo, H. H. (2021). Role of artificial intelligence applications in real-life clinical practice: Systematic review. Journal of Medical Internet Research, 23(4), e25759. https://doi.org/10.2196/25759

Yusuf, A., Bello, S., Pervin, N., & Tukur, A. K. (2024). Implementing a proposed framework for enhancing critical thinking skills in synthesizing AI-generated texts. Thinking Skills and Creativity, 53, 101619. https://doi.org/10.1016/J.TSC.2024.101619

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? In International Journal of Educational Technology in Higher Education (Vol. 16, Issue 1). Springer Netherlands. https://doi.org/10.1186/s41239-019-0171-0

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

Zhang, S., Zhao, X., Zhou, T., & Kim, J. H. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. International Journal of Educational Technology in Higher Education, 21(1). https://doi.org/10.1186/s41239-024-00467-0

Zheng, K., Shen, Z., Chen, Z., Che, C., & Zhu, H. (2024). Application of AI-empowered scenario-based simulation teaching mode in cardiovascular disease education. BMC Medical Education, 24(1). https://doi.org/10.1186/s12909-024-05977-z

Zou, X., Su, P., Li, L., & Fu, P. (2024). AI-generated content tools and students' critical thinking: Insights from a Chinese university. IFLA Journal, 50(2), 228–241. https://doi.org/10.1177/03400352231214963

Tables 1.
Results of studies included in this SR.

Academic field	Degree	AI type	Activities in which AI has been used	Pretest-Posttest	Critical Thinking measurement tool	Duration of the study	Main conclusion	Study
Arts and Humanities	Foreign language	Chat GPT	Metaphors about GenAI in learning English	No	Self-reported (students' perception and attitudes towards AI)	-	Promotes optimal learning. Potential loss of critical thinking skills.	(Yan et al., 2024)
			Collaborative argumentation, guiding students in stages of preparation, interaction, and reflection	Yes	Scale Likert modified (Chai et al., 2015)	6 weeks	Improves awareness of critical thinking, reasoned speaking and collaboration.	(Darmawansah et al., 2025)
		Quillbot	Activities of reviews and text modification	Yes	California Critical Thinking Test (CCTST) (Facione et al., 1994)	56 weeks	Critical thinking and self-directed learning help students use AI responsibly and effectively when writing in English.	(Xiaolei & Teng, 2024)
		Automatic Writing Evaluation (AWE)	AI-powered writing and feedback processes; Task assignment; Initial composition; Automatic submission and feedback; AI evaluation and suggestions; Detailed revision suggestions; Revision and reflection	Yes	Critical thinking questionnaire (Lin et al., 2019)	4 weeks	Improves critical thinking.	(Liu et al., 2023)
	Philosophy	Chat GPT	Student chat logs on ChatGPT	No	Students' self-reported perception. Students' experiences and perceptions regarding critical thinking in an AI-enhanced environment.	16 weeks	Strengthened critical thinking as part of faster, higher-quality research progress, improved academic and professional skills, greater confidence, and closer supervisory relationships.	(Dai et al., 2023)
Science	Physics	Chat GPT	"Spot the Bot". To identify errors in AI-generated responses and reflected on their understanding of scientific concepts.	No	The impact of responses on the perception of scientific accuracy and linguistic quality.	15-20 min	Students with sufficient prior knowledge can adequately evaluate scientific quality, encouraging reflection and critical thinking than students with less prior knowledge.	(Dahlkemper et al., 2023)
	Science	Chat GPT	Academic help, writing support, and exploring new information while also using it to satisfy curiosity, address personal issues, pass time, and occasionally seek emotional support.	Does not specify	Students' self-reported perception.	4 weeks	Negative effects on critical thinking and a decline in personal skills.	(Zhang et al., 2024)
Health Sciences	Medicine	AI-based clinical decision support systems	Project design and resolution of specific medical problems	No	Qualitative evaluation of deliverables.	48 weeks	Improves motivation and development of digital skills and critical thinking.	(Tsopra et al., 2023)
		Chat GPT	Clinical diagnosis after random assignment	Yes	Preliminary questionnaire, quality of the product, case-by-case questionnaire, and final questionnaire.	8-16 weeks	Students' critical thinking (in terms of information search, number of diagnostic hypotheses considered, diagnostic accuracy, and confidence level) may be impaired without prior training.	(Kämmer et al., 2024)

		Chat GPT	Scenario simulation	Yes	Clinical Critical Thinking Scale based on Ennis´s critical thinking framework (Ennis, 1996).	48-96 weeks	Improves critical thinking skills (higher theoretical knowledge score, clinical skills performance, and clinical critical thinking)	(Zheng et al., 2024)
	Faculty of Medicine	Chat GPT	Perceptions about the use of ChatGPT	No	Students' self-reported perception. 25 questions (open and closed-ended) about attitude towards AI.	No	Encourages critical thinking, as students are forced to question, verify, and contrast the information generated to counteract potential errors or fabrications.	(Araujo & Cruz-Correia, 2024)
	Physical therapy	Chat GPT	Response generation, reflection on learning, reviewing queries, evaluating responses.	Does not specify	Does not specify.	Does not specify	Synthesizes information efficiently but students need time to ensure that the protocol is correct and that fact-checking is necessary.	(Thompson & Fecher, 2024)
	Odontology	AiDental	Automated prosthesis design.	Yes	Qualitative evaluation of deliverable. Both quantitative and qualitative analyses.	2.5 months	Improves decision-making and critical thinking skills. More effective among younger people.	(Mahrous et al., 2023)
	Physiology	ChatGPT Bing	Creation of schedules, materials, providing feedback on students' notes, identifying gaps and areas that need further study, searching for additional resources, and exam preparation.	No	Does not specify.	-	Supplemental resource to amplify critical thinking.	(Favero, 2024)
	Nursing	Chat GPT	Learning activities.	Yes	Critical Thinking Scale (Lin et al., 2019) of the Critical Thinking Disposition Questionnaire (Chai et al., 2015)	4 hours	Improves students' critical thinking skills, problem-solving, and learning enjoyment	(Chang et al., 2024)
		Chat GPT	Creating and analysing case scenarios for NCLEX administration using the NCSBN Clinical Judgment Measurement Model (NCJMM).	No	Qualitative analysis of deliverables.	-	Improves critical thinking and prepares for an increasingly technology-dependent healthcare environment.	(Simms, 2024)
		IntelliMetric	Academic writing tasks.	Yes	Qualitative analysis of deliverables.	14 weeks	Can help assess students' writing performance and improve their writing skills (linked to critical thinking).	(Wolf & Wolf, 2023)
		Chat GPT	AI opinion questionnaire.	No	Students' self-reported perception.	Does not specify	Perceived as attractive and potentially beneficial, although 31.3% of respondents are concerned that it "limits opportunities to use critical thinking."	(Bonacaro et al., 2024)

Social and Legal Sciences	Business	Chat GPT	Proofreading and Language Correction.	Yes	Bloom's taxonomy (Krathwohl, 2002) . Quantitative and qualitative tests (multiple-choice questions and Likert scales, open-ended questions to delve deeper into students' experiences)	Does not explicitly state the total duration of the study	Could improve both basic and advanced levels of critical thinking, although it is a complex process influenced by several factors.	(Essien et al., 2024)
	International relations	Chat GPT	Create the SWOT (Strengths, Weaknesses, Opportunities, Threats) matrix and the MICMAC (Multiplied Cross-Impact Matrix Applied to a Classification) matrix.	Yes	Qualitative analysis of deliverables from the student activities.	4 weeks	Strengthening your communication, critical thinking, and logical reasoning skills.	(Michalon & Camacho-Zuñiga, 2023)
Engineering and Architecture	Computer Sciences	AI-assisted teaching platform	Learning activities (problem solving), student monitoring and evaluation.	Yes	Qualitative analysis of deliverables from the student activities (argumentation skills and the development of critical thinking skills).	24 weeks	Improves critical thinking, speculative ability, motivation and autonomous learning.	(He et al., 2023)
		Scrum IA	Problem solving	Yes	Students' self-reported perception.	10 weeks	Problem-solving skills, critical thinking and communication skills, teamwork skills, and the ability to work with businesses	(Augusto Menten de Barros et al., 2023)
	Computer Sciences	NLP models: BERT GPT-2	Generation of personalized materials in Python.	Yes	MSLQ learning strategy subscale (Pintrich et al., 1991; Pintrich et al., 1993)	8 weeks	Improves learning performance, critical thinking, metacognitive self-regulation, and effort regulation.	(Huang et al., 2023)
	Technology	Chat GPT	Integrated STEM instructional designs working collaboratively with ChatGPT: planning, problem-solving, and pedagogical development	Yes	Modified version of (Chen, 2000) questionnaire. Based on the California Critical Thinking Disposition Inventory (CCTDI) (N. C. Facione et al., 1994)	8 weeks	Promotes more systematic critical thinking, but human teachers can provide advantages for the final product.	(Li et al., 2024)
	Information technologies	PSoC 6 microcontroller technology	Detection of musical notes, voice and background noise.	Yes	Qualitative analysis of deliverables.	Does not specify	Positive attitude towards the use of AI and the importance of critical thinking when using it.	(Modran et al., 2024)
	Engineering	Chat GPT	Problem solving in a mathematical context, scheduled tasks outside the classroom and the learning process in general.	No	Students' self-reported perception of AI influence.		Adaptation of teaching strategies is needed to ensure the development of critical thinking	(Sánchez-Ruiz et al., 2023)
	Electrical and mechanical engineering	Chat GPT	Preparation of laboratory reports.	Yes	Inclusive assessment rubric adapted from the 2014 Writing Program Administrators Outcomes Statement to evaluate critical thinking and composing.	2 weeks	Enables improvements in critical thinking, reading, and composition, although it does not necessarily foster profound improvements in critical thinking as it can lead to the inclusion of incorrect or superficial information.	(Kim et al., 2024)
Education	Education for primary school/ Teaching Master	Makeblock	Screen-free programming using computational thinking fundamentals.	Yes	Evaluating the adoption of STEAM tools, changes in teacher beliefs, and student engagement through pre/post-tests and interviews.	7 weeks	Positive attitude toward the use of AI, applying critical thinking to its use.	(Henze et al., 2022)

	Educational /teaching	Chat GPT Gen IA	Interactive analyses: PESTLE analysis. Porter's Five Forces analysis. AI-generated cases: To present real-world business scenarios. Ethical dilemma resolution: Through an ethical dilemma chatbot.	No	Qualitative analysis of deliverables from the student activities. Students' self-reported perception.	24 weeks	Can diminish students' critical thinking. Educators should implement strategies that encourage independent analysis and promote reasoning in their decisions.	(Hyde et al., 2024)
	Education for primary school/ Teaching Master	Hereinafter IA	Addressing "deepfakes" in arts education to develop students' critical thinking and visual literacy.	Does not specify	Semi-structured questionnaire following Bloom's taxonomy.	16 weeks	The development of critical thinking is essential for the autonomy of thought, both in their creative and misinformative aspects.	(Murillo-Ligorred et al., 2023)
	Teaching	ChatGPT Midjourney	Story creation.	Yes	Turkish adaptation of Reflective Thinking Test. (Kember et al., 2000) by Başol & Evin-Gencel, (2013).	5 weeks	Can contribute to critical thinking, but there are no significant differences.	(Saritepeci & Yildiz Durak, 2024)
Mixed	Arts and Music	Chat GPT Augmented reality	Learning tasks to help students.	Yes	Computational thinking questionnaire (adapted from Korkmaz et al., 2017)	14 days	Significantly greater improvement in their computational thinking skills compared to the control group, including creativity, logical computation, critical thinking, and problem-solving skills	(P. H. Lin & Chen, 2020)
	Environmental Sciences, Biology, Chemistry, Geography and Physics, Nanotechnology, Humanities or Language	Chat GPT	Conduct a seminar on a scientific article or report. Identify a research question or main topic and formulate a ChatGPT task. Validate the answer using the original source and other sources	No	Qualitative analysis of deliverables.	48 weeks	Improves critical thinking skills.	(Šedlbauer et al., 2024)
	Engineering, humanities and social sciences, natural sciences, business and economics, and other fields	ChatGPT DALL-E Bing	Investigates the use of AI-generated content tools (reasons for use, perceived benefits, risks and limitations, and the importance of critical thinking skills for future professional development)	No	Student's self-reported measurements.	No	After using AI, Engineering had the highest percentage of critical thinking (62.98%), folLowed by humanities and social sciences (13.28%), natural sciences (8.7%), business and economics (3.06%) and other fields (11.99%).	(Zou et al., 2024)
	Different disciplinary areas	AI platform developed to assess complex thinking	Work on an interactive platform through 4 modules on "sharing economy".	Yes	"eComplexity" and "Diapason" instruments.	6 hours	Increase in the scientific thinking sub competence.	(Sanabria-Z et al., 2023)

	Course on the historical and philosophical foundations of adult and community education and exploration of the nature and scope of the field	Chat GPT	Asynchronous discussions in ChatGPT, critique and reflect on information generated by AI.	Yes	Critical Thinking Scale, adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991) Course Experience Questionnaire (CEQ) developed by Wilson et al. (1997).	24 weeks	Encourage constructive conversations and develop essential skills, such as critical thinking and knowledge exploration.	(X. Lin et al., 2024)
	Computer Science, Islamic Education management, and other programs	Mitsuku Replika	Interaction with "AI friends".	Yes	AI-related questionnaire to measure students' critical thinking propensity and reactions to AI (Miri et al., 2007).	12 weeks	Integrating AI into English teaching significantly improves critical thinking skills and language proficiency.	(Muthmainnah et al., 2022)
	Chemistry and sports and health	Chat GPT	Interaction with chat GPT.	Yes	Questionnaire that includes specific elements related to decision-making, Surveys to identify assessment competency factors and analyse query behaviour using ChatGPT.	-	The use of AI tools can lead to a dependency that affects critical thinking skills.	(Dindorf et al., 2024)
	Unspecified	Chat GPT	Research and learning.	No	Does not specify.	4 weeks	It is vital to develop assessments that support responsible use of ChatGPT, maintaining students' critical thinking skills and originality in writing.	(Shahzad et al., 2024)
	Unspecified	ChatGPT (or similar)	Activities to familiarize, conceptualize, inquire, evaluate, synthesize.	Yes	Short form of the Critical Thinking Self-Assessment Scale (CTSAS) psychometric inventory (Payan-Carreira et al., 2022).	Study training phase: 6 hours. The total duration of the research project is not explicitly specified.	Improves critical thinking skills by analysing AI-generated texts.	(Yusuf et al., 2024)
	Unspecified	Chat GPT	Drawing analysis.	Yes	Critical Thinking Tendency Questionnaire (Chai et al., 2015).	6 hours	Students with a high growth mindset have greater critical thinking and use AI more effectively.	(Tu, 2024)

Tabla 2.
Description of the type of activity, number of studies addressing each activity, and percentage of publications classified by activity type.

Type	Nº articles (N=41)	%	Study
Text generation	20	48,78%	(Darmawansah et al., 2025; Favero, 2024; He et al., 2023; Hyde et al., 2024; Kim et al., 2024; Li et al., 2024; X. Lin et al., 2024; Mahrous et al., 2023; Michalon & Camacho-Zuñiga, 2023; Modran et al., 2024; Q Huang et al., 2023; Šedlbauer et al., 2024; Shahzad et al., 2024; Simms, 2024; Tu, 2024; Wolf & Wolf, 2023; Xiaolei & Teng, 2024; Yan et al., 2024; Yusuf et al., 2024; Zou et al., 2024)
Writing/performance assessment	19	46,34%	(Araujo & Cruz-Correia, 2024; Bonacaro et al., 2024; Chang et al., 2024; Dahlkemper et al., 2023; Dai et al., 2023; He et al., 2023; Li et al., 2024; Liu et al., 2023; Mahrous et al., 2023; Modran et al., 2024; Q Huang et al., 2023; Sanabria-Z et al., 2023; Saritepeci & Yildiz Durak, 2024; Šedlbauer et al., 2024; Thompson & Fecher, 2024; Tu, 2024; Wolf & Wolf, 2023; Yusuf et al., 2024; Zheng et al., 2024)
Problem solving	6	14,63%	(Menten de Barros et al., 2023; Dindorf et al., 2024; Hyde et al., 2024; Kämmer et al., 2024; Sánchez-Ruiz et al., 2023; Tsopra et al., 2023)
Questioning/chatbot	3	7,31%	(Hyde et al., 2024; Muthmainnah et al., 2022; Yan et al., 2024)
Augmented reality	1	2,43%	(Lin & Chen, 2020)
Research assistance	1	2,43%	(Shahzad et al., 2024)
Not specified	4	9,75%	(Essien et al., 2024; Henze et al., 2022; Murillo-Ligorred et al., 2023; Zhang et al., 2024)

In this table, we describe the type of activity (problem solving, writing/performance assessment, text generation, questioning/chatbot, image generation, Augmented reality, research assistance and not specified), the number of articles in which each activity is developed, and the percentage of publications classified by activity type

Table 3.
Summary risk of bias in individual studies

Study	Bias selection (Random sequence)	Bias selection (Allocation concealment)	Performance bias	Detection bias	Attrition bias	Reporting bias	Automation bias	Representative data	Transparency / Reproducibility
(P. H. Lin & Chen, 2020)	Unclear	Unclear	Unclear	Unclear	High	High	Unclear	Unclear	Unclear
(Henze et al., 2022)	Unclear	Unclear	Unclear	Unclear	High	High	High	Unclear	Low
(Muthmainnah et al., 2022)	Unclear	High	Unclear	Unclear	High	High	Unclear	Unclear	Unclear
(Dahlkemper et al., 2023)	High	Low	High	High	High	Low	Low	Unclear	Low/Unclear
(Dai et al., 2023)	Unclear	Unclear	Unclear	High	High	High	Low	Unclear	Unclear
(He et al., 2023)	Low	High	Unclear	Unclear	Low	Low	Low	High	Low/High
(Liu et al., 2023)	High	Unclear	Unclear	Unclear	High	High	Low	High	Unclear
(Mahrous et al., 2023)	Low	Unclear	Low	Low	Low	Unclear	Low	High	Low/High
(Augusto Menten de Barros et al., 2023)	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	High	Unclear	Unclear/High
(Michalon & Camacho-Zuñiga, 2023)	Doesn't apply	Doesn't apply	Unclear	Unclear	Low	Unclear	Low	Unclear	Unclear
(Murillo-Ligorred et al., 2023)	Doesn't apply	Unclear	Unclear	Unclear	Low	Low	Low	High	Low/High
(Q Huang et al., 2023)	Low	High	Unclear	Unclear	Low	Low	Unclear	High	Low/High
(Sanabria-Z et al., 2023)	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Unclear	Unclear

(Sánchez-Ruiz et al., 2023)	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	High	Low/High
(Tsopra et al., 2023)	High	High	Low	Low	Low	Low	Unclear	Unclear	Low/High
(Wolf & Wolf, 2023)	High	High	High	High	Low	Low	Low	High	Low/High
(Araujo & Cruz-Correia, 2024)	Low	Doesn't apply	High	High	Unclear	Unclear	Low	Unclear	Unclear
(Bonacaro et al., 2024)	Unclear	Unclear	Unclear	Unclear	High	High	High	High	Unclear
(Chang et al., 2024)	Low	High	Unclear	Unclear	High	High	Unclear	High	Unclear
(Dindorf et al., 2024)	High	Unclear	Low	Low	Low	Low	Low	High	Unclear/High
(Essien et al., 2024)	Low	High	High	Low	Unclear	Unclear	Low	Unclear	Unclear
(Favero, 2024)	Doesn't apply	Doesn't apply	Doesn't apply	Doesn't apply	High	High	Low	High	High
(Hyde et al., 2024)	High	Unclear	Unclear	Unclear	Low	Low	Low	Unclear	Low
(Kämmer et al., 2024)	Low	Low	Low	Low	Low	Low	Low	Unclear	Low/Unclear
(Kim et al., 2024)	High	High	Low	High	High	High	Low	Unclear	Unclear
(Li et al., 2024)	Low	Unclear	Low	Low	High	High	Low	High	High
(X. Lin et al., 2024)	Unclear	Unclear	Unclear	Unclear	Low	Low/Unclear	Low	High	Unclear
(Modran et al., 2024)	High	Doesn't apply	Doesn't apply	Doesn't apply	High	High	Low	High	Low
(Saritepeci & Yildiz Durak, 2024)	Low	Unclear	Unclear	Unclear	Low	Low	Low	Unclear	Low
(Šedlbauer et al., 2024)	High	High	High	High	Low	Low	High	High	Low/High

(Shahzad et al., 2024)	High	Doesn't apply	High	High	Low	Low	Unclear	High	Unclear/High
(Simms, 2024)	High	High	High	High	High	High	Low	Unclear	Unclear
(Thompson & Fecher, 2024)	High	Unclear	Unclear	Unclear	High	High	Low	Unclear	Low/Unclear
(Tu, 2024)	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Unclear	Low/Unclear
(Xiaolei & Teng, 2024)	Low	Unclear	Unclear	Unclear	High	High	Unclear	Unclear	Unclear
(Yan et al., 2024)	Unclear	Unclear	Unclear	Low	Low	Low	Low	High	Low/Unclear
(Yusuf et al., 2024)	Low	Unclear	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Low
(Zhang et al., 2024)	Unclear	Doesn't apply	Doesn't apply	Doesn't apply	High	High	Low	High	Low
(Zheng et al., 2024)	Low	Unclear	Unclear	Unclear	Low	Unclear	Low	High	Low/High
(Zou et al., 2024)	High	Doesn't apply	High	High	Unclear	Unclear	Low	High	Low/High
(Darmawansah et al., 2025)	High	High	Unclear	Unclear	Low	Low	Low	Unclear	Low/High
High	3	1	1	1	3	3	0	4	0
Low	3	0	0	1	4	3	8	0	2
Unclear	3	6	7	6	2	3	1	5	1

Author's judgements about all type of bias for each publication reviewed. Selection bias (random sequence generation and allocation concealment), performance bias (knowledge about the use of AI and blinding of assessors), detection bias (random outcome assessment and blinding), attrition bias (incomplete outcome data), reporting bias (selective outcome reporting) and additional sources of bias (consideration of errors contributed by AI, representative data, transparency, and reproducibility).