MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2026, 22(12), em2961
ISSN:1305-8223 (online)
OPEN ACCESS Literature Review https://doi.org/10.29333/ejmste/19893

How students engage with LLM-based GenAI to develop computational thinking: A systematic review

Mustafa Mustafa [1], Tatang Herman [1*], Yaya S. Kusumah [1], Nurjanah Nurjanah [1], Trysa Gustya Manda [1]

[1] Department of Mathematics Education, Indonesia University of Education, Bandung, INDONESIA

Received 11 June 2026 ▪ Accepted 13 September 2026

Abstract

This systematic literature review examines how students engage with large language model-based generative artificial intelligence (GenAI) in developing computational thinking (CT). Following PRISMA 2020 procedures, 14 of 212 peer-reviewed studies published between 2021 and 2026 were synthesized to investigate how GenAI functions along a continuum from problem-solving tool to thinking partner. The review identifies four student-AI interaction patterns: surface, scaffolded, achiever-linked, and dialogic. Students with stronger CT proficiency and AI literacy engaged GenAI dialogically through critique, verification, and reflective prompting, whereas lower-proficiency learners relied on answer-oriented interactions linked to substitutive offloading and weaker problem-solving. GenAI supported algorithmic thinking, creativity, and reflection when students retained task agency, but risked over-reliance otherwise. Most studies positioned GenAI at a cognitive tool mid-point, yet sustained CT development depended on designs promoting epistemic agency and reflective dialogue. The review contributes a continuum framework and seven pedagogical safeguards that shift AI use toward thinking partnership.

Keywords: computational thinking, LLM-based GenAI, pedagogical safeguards, student-AI interaction, thinking partner, systematic review

INTRODUCTION

The rapid public adoption of large language model-based generative artificial intelligence (LLM-based GenAI), particularly after the release of ChatGPT in November 2022, has created a decisive shift in educational practice. Conversational systems such as ChatGPT, Gemini, Claude, and Copilot are no longer peripheral technologies; they are increasingly used by students to access information, generate ideas, summarize texts, write code, solve problems, and refine academic work. Unlike earlier rule-based tutoring systems, LLM-based GenAI provides natural-language dialogue, explanation generation, artefact production, critique, and iterative feedback, making it continuously available as a linguistic interlocutor in students' learning processes (Arum et al., 2025; Rasul et al., 2023; Strzelecki, 2023). This development raises a central educational question: does GenAI strengthen students' thinking, or does it risk replacing the cognitive work that students should perform themselves?

This question is especially important for computational thinking (CT). Since CT involves problem formulation, decomposition, abstraction, algorithmic reasoning, pattern recognition, debugging, and iterative refinement (Shute et al., 2017; Wing, 2006), it is closely connected to the way students communicate with and evaluate outputs from LLM-based GenAI. In the GenAI era, students are not only required to produce answers, but also to frame problems clearly, interrogate machine-generated responses, revise prompts, verify explanations, and decide whether AI outputs should be accepted, modified, or rejected (Hsu, 2025; Sánchez Muñoz et al., 2025). Thus, CT remains highly relevant, not merely as a programming-related competence, but as a broader cognitive foundation for productive human-AI interaction (Hsu, 2025; Shute et al., 2017; Wing, 2006).

© 2026 by the authors; licensee Modestum. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).
✉ mustafa@upi.edu ✉ tatangherman@upi.edu **(*Correspondence)** ✉ yskusumah@upi.edu ✉ nurjanah@upi.edu
✉ trysagustya@upi.edu

Contribution to the literature

- This review reframes the study of GenAI in CT from whether the technology improves outcomes to how students interact with it, introducing a problem-solver-to-thinking-partner continuum as an analytical lens.
- It identifies four recurring student-AI interaction patterns and links them to learners' CT proficiency and AI literacy, clarifying empirical tensions in prior studies that report both gains and erosion of problem-solving.
- It synthesizes seven pedagogical safeguards mapped to the continuum, offering educators and researchers an actionable basis for designing interaction architectures that move GenAI use from substitutive problem-solving toward dialogic thinking partnership.

Recent studies over the last five years show that LLM-based GenAI and related AI tools can support CT development across diverse educational contexts, including programming, mathematics, information technology, robotics, design thinking, and interdisciplinary learning. Structured GenAI use has been associated with improvements in algorithmic thinking, creativity, problem refinement, reflective thinking, engagement, and learning performance (Ameen et al., 2024; Gong et al., 2025; Hijón-Neira et al., 2024; Liao et al., 2024). Beyond GenAI, a meta-analysis of science-education interventions found a moderate positive effect on CT, indicating that well-designed instruction can strengthen CT (Kıyıcı & Yamak, 2026). Other studies also suggest that GenAI can support inquiry, scaffolding, feedback, and co-creation when learning activities are pedagogically designed rather than left entirely to students' spontaneous use (Bagdonaitė & Dagienė, 2025; Otto et al., 2025; Sánchez Muñoz et al., 2025).

At the same time, the literature also documents substantial risks. When students use GenAI mainly to obtain direct answers, complete assignments, or bypass reasoning, the technology may encourage cognitive offloading, over-reliance, shallow engagement, weaker debugging, and reduced independent problem-solving (Gerlich, 2025; Grassini, 2025). These risks are particularly visible when learners have limited prior knowledge, weak AI literacy, or insufficient pedagogical guidance (Bastani et al., 2024; Gerlich, 2025; Grassini, 2025; Yusuf et al., 2024). The critical issue, therefore, is not simply whether LLM-based GenAI is beneficial or harmful, but how students interact with it during CT learning.

This review approaches that issue through a continuum from problem solver to thinking partner. In problem-solver mode, students delegate cognitive work to LLM-based GenAI in order to obtain completed answers with minimal critique, refinement, or metacognitive engagement. In thinking-partner mode, students engage GenAI dialogically to test, critique, revise, and extend their own reasoning while retaining epistemic agency over the task. These modes are not fixed properties of the tool itself. The same GenAI system may occupy any of the three continuum positions, functioning as a problem solver, a cognitive tool, or a thinking partner, depending on task design, student proficiency, prompt strategy, and pedagogical support (Gong et al., 2025; Hsu, 2025; Sánchez Muñoz et al., 2025). The notion of AI as a thought or thinking partner derives from prior work in cognitive science and human-AI interaction (Salomon et al., 1991); the original contribution here is not the term but its organization into a three-position continuum and its systematic application to CT. Each factor operates through a distinct route: task design sets the cognitive demand that the learner must retain; student proficiency and prior knowledge determine whether the learner can evaluate and verify AI output; prompt strategy governs whether the exchange is single-shot or iterative and dialogic; and pedagogical support regulates how far the AI is permitted to substitute for, rather than scaffold, the learner's reasoning.

Although the growing literature on AI in education has examined learning outcomes, student perceptions, programming education, and AI-supported instruction (Bagdonaitė & Dagienė, 2025; Huang & Tseng, 2026; Otto et al., 2025), a thematic reading of these reviews reveals three gaps. First, prior reviews rarely isolate how LLM-based GenAI is used by students specifically for CT development, typically treating CT as one outcome among many rather than as the focal construct (Bagdonaitė & Dagienė, 2025; Otto et al., 2025). Second, existing studies often emphasize effectiveness outcomes while giving less attention to the student-AI interaction patterns through which those outcomes emerge (Huang & Tseng, 2026; Otto et al., 2025). Third, the field still offers limited guidance on pedagogical safeguards that can prevent GenAI from becoming a cognitive shortcut that weakens CT (Gerlich, 2025; Grassini, 2025). **Table 1** summarizes these substantive, analytical, and pedagogical gaps and links them directly to the research questions (RQs) of this review.

Guided by these gaps, this systematic literature review (SLR) addresses three RQs:

RQ1: How is LLM-based GenAI used by students along a continuum from problem-solving tool to thinking partner in CT learning?

Table 1. Research gaps addressed by the review and their linkage to the RQs

Type of gap	What existing literature does	What remains underexplored	Linked RQ	Key sources
Substantive	Examines AI in education broadly and reports effects on achievement, engagement, and attitudes.	Rarely isolates how LLM-based GenAI is positioned and used by students specifically for computational thinking development.	**RQ1**	Bagdonaitė and Dagienė (2025), Otto et al. (2025)
Analytical	Focuses on effectiveness outcomes and student perceptions of GenAI.	Does not systematically characterize student–AI interaction patterns or relate them to learners' CT proficiency or prior experience.	**RQ2**	Huang and Tseng (2026), Otto et al. (2025)
Pedagogical	Advocates AI as a support tool and reports strong enthusiasm among educators and students.	Offers limited guidance on pedagogical safeguards needed to prevent GenAI from becoming a cognitive shortcut that weakens CT.	**RQ3**	Gerlich (2025), Grassini (2025)

RQ2: What student-AI interaction patterns are reported during CT learning with LLM-based GenAI, and how are these patterns related, first, to learners' CT proficiency and, second, to their prior experience?

RQ3: what cognitive benefits, risks, and pedagogical safeguards are reported regarding student interaction with LLM-based GenAI in CT development?

The novelty of this review lies in shifting the analytical focus from whether LLM-based GenAI improves learning outcomes to how students interact with GenAI in developing CT. Unlike prior reviews that mainly examine AI in education broadly (Bagdonaitė & Dagienė, 2025), achievement effects (Otto et al., 2025), or GenAI use in programming (Huang & Tseng, 2026), this review conceptualizes student-GenAI interaction along a problem solver to thinking partner continuum. It contributes an interaction-centered synthesis of LLM-based GenAI in CT learning, identifies patterns of student-AI engagement linked to CT proficiency and prior experience, and synthesizes pedagogical safeguards intended to strengthen rather than substitute for students' CT. A recent review that jointly addresses AI, robotics, and CT similarly reports that integrated treatments of these areas remain scarce, underscoring the need for focused syntheses such as the present one (Sendarrubias-Hipólito et al., 2026).

LITERATURE REVIEW

This framework integrates CT theory, classical cognitive-tool perspectives, and contemporary human-GenAI interaction research to ground the problem-solver-to-thinking-partner continuum that organizes the review.

Computational Thinking in the LLM-Based GenAI Era

CT has been defined as a way humans think and, more specifically, as the formulation of problems and solutions for an information-processing agent (Wing, 2006, 2008). In educational research, CT is commonly conceptualized through three complementary lenses. First, it is treated as a cognitive process involving decomposition, abstraction, and algorithmic reasoning (Shute et al., 2017; Wing, 2006). Second, it is understood as an educational outcome reflected in measurable competencies such as creativity, algorithmic thinking, critical thinking, and problem-solving (Korkmaz et al., 2017). Validated instruments for secondary-school science and mathematics students provide further support for treating CT as a measurable competence (Rohaeti & Huda, 2025). Third, in LLM-based GenAI environments, CT can be interpreted as an interactional practice that appears in prompt-based moves such as decomposing prompts, refining outputs, and debugging AI responses (Hsu, 2025). Earlier syntheses of computational-thinking research likewise document the breadth of these CT sub-skills across problem-solving contexts (Ilic et al., 2018).

For synthesis, this review uses five CT dimensions: decomposition, pattern recognition, abstraction, algorithmic thinking, and automation/debugging. These five dimensions derive from the foundational, technology-independent accounts of CT (Brennan & Resnick, 2012; Shute et al., 2017; Wing, 2006); and align with the CT Scale (Korkmaz et al., 2017); they are used here because they map closely onto observable student-GenAI interaction moves. Decomposition appears when students split a complex task into a planned sequence of prompts instead of asking for a complete solution at once (Hsu, 2025; Liao et al., 2024). Pattern recognition is reflected in students' ability to identify recurring structures across AI outputs, reuse successful prompt patterns, and derive transferable solution templates (Hsu, 2025). Abstraction is visible when students design concise and explicit prompts that remove contextual noise and define the AI's task at an appropriate level of generality (Gong et al., 2025; Hsu, 2025). Algorithmic thinking is expressed through logical and stepwise prompting, the specification of constraints and expected output structures, and the sequencing of prompts that build on prior results (Gong et al., 2025; Hijón-Neira et al., 2024; Korkmaz et al., 2017; Shute et al., 2017). Automation/debugging appears when students iteratively refine prompts, critique AI outputs against

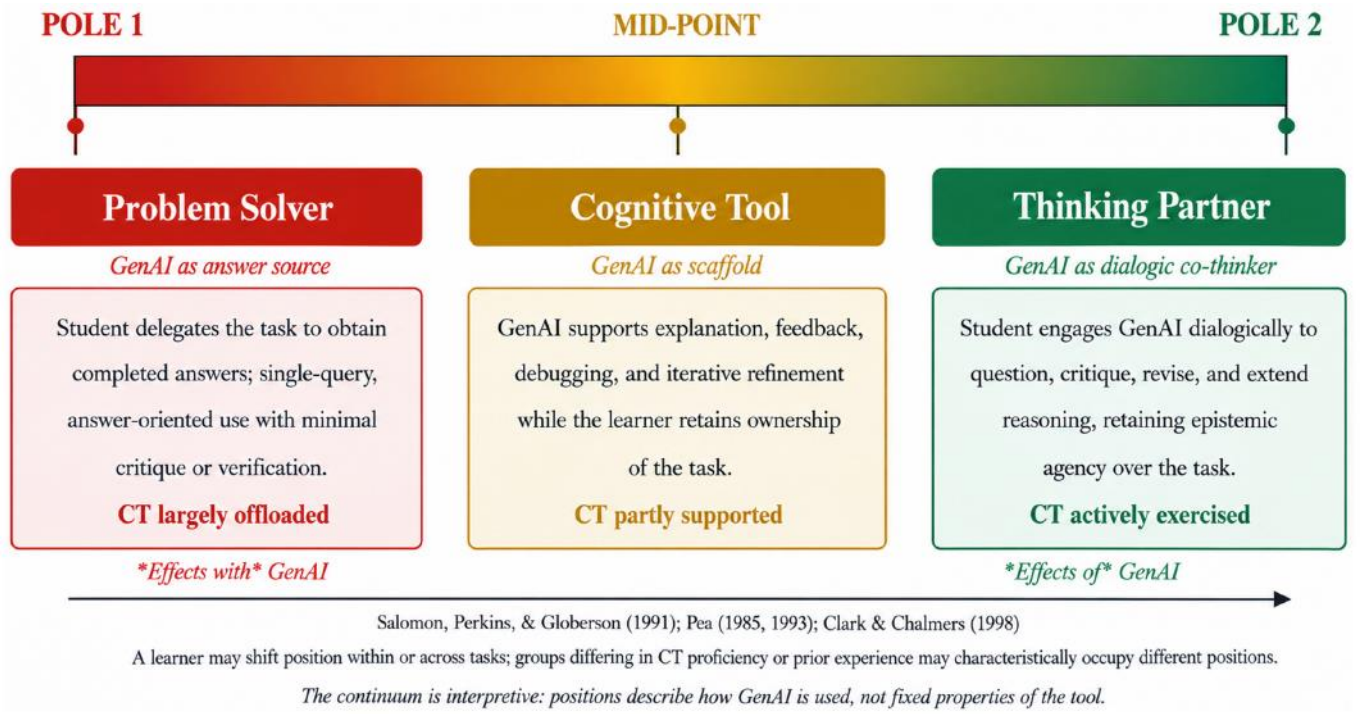


Figure 1. The student-LLM-based GenAI interaction continuum in CT development (the author's own elaboration)

expected outcomes, and use targeted re-prompts to correct errors or address gaps (Brennan & Resnick, 2012; Hsu, 2025; Liao et al., 2024; Shute et al., 2017). Because LLMs respond to natural-language instructions, CT in the GenAI era functions as a literacy for communicating with probabilistic information-processing agents (Hijón-Neira et al., 2024; Hsu, 2025; Wing, 2008).

Theoretical Anchors and Cognitive Offloading

Three theoretical contributions clarify the continuum used in this review. Salomon et al. (1991) distinguished effects with a technology, namely performance gains observed while using a tool, from effects of a technology, namely enduring cognitive residues that remain after the tool is withdrawn. Pea (1985, 1993) distinguished technologies that amplify existing cognition from technologies that reorganize mental functioning. Clark and Chalmers' (1998) extended-mind hypothesis is used here in a moderate and interactional sense: when student-AI dialogue is sustained, evaluative, and integrated through revision, AI can function as a cognitive partner within the learning interaction (Sánchez Muñoz et al., 2025). Together, these perspectives explain why LLM-based GenAI may either accelerate answer production or contribute to the restructuring of CT practice (Bagdonaitė & Dagienė, 2025; Hsu, 2025; Otto et al., 2025).

This theoretical lineage also clarifies the role of cognitive offloading. Productive offloading reduces working-memory burden while preserving the student's epistemic agency. Substitutive offloading, by contrast, transfers cognitive responsibility to AI and is associated with weaker critical thinking and problem-solving (Bastani et al., 2024; Gerlich, 2025; Grassini, 2025; Liao et al., 2024). This distinction is central to the review because it helps interpret whether LLM-based GenAI supports students' CT development or replaces the reasoning processes that CT instruction is intended to cultivate.

The Problem Solver to Thinking Partner Continuum

The continuum is organized around three anchor positions: the problem solver pole, the cognitive tool mid-point, and the thinking partner pole. **Figure 1** presents the continuum, its theoretical anchors, and its observable indicators. In this framework, thinking partner is treated as an interactional construct rather than an anthropomorphic claim about AI.

At the problem solver pole, students outsource cognitive work through single-query requests, verbatim use of AI outputs, and limited critique. These interactions primarily produce effects with the technology and instantiate substitutive offloading (Bastani et al., 2024; Gong et al., 2025; Grassini, 2025; Salomon et al., 1991). At the cognitive tool mid-point, GenAI scaffolds reasoning through explanations, worked examples, corrections, and stepwise refinement. In this position, students remain owners of the task, and offloading is largely productive (Gong et al., 2025; Hijón-Neira et al., 2024; Liao et al., 2024; Otto et al., 2025). At the thinking partner pole, students co-construct thinking through critique, metacognitive prompting, debating, synthesis across turns, and transfer to new problems. Epistemic vigilance, understood as the disposition to question, verify, and revise AI claims, is a constitutive marker of this pole (Gong et al., 2025; Hsu, 2025; Sánchez Muñoz et al., 2025).

The continuum is descriptive and pedagogically actionable. The same LLM-based GenAI system can occupy different positions depending on task design, prompt strategy, student proficiency, AI literacy, and pedagogical scaffolding. Two forms of variation should be distinguished: within a given learner, the position may shift across the phases of a single task or across tasks; across learners, groups differing in CT proficiency or prior experience may characteristically occupy different positions. The latter reflects between-group differences observed cross-sectionally rather than change over time. Instructional design can move interaction toward the thinking partner pole through prompt literacy, reflection, graduated support, and critical-evaluation tasks (Bagdonaitė & Dagienė, 2025; Hsu, 2025; Sánchez Muñoz et al., 2025; Wu & Yang, 2022). This framework guides the review's analysis of AI positioning in **RQ1**, student-AI interaction patterns and CT-proficiency linkages in **RQ2**, and cognitive benefits, risks, and pedagogical safeguards in **RQ3**. The operational coding framework that translates the continuum into extractable categories is detailed in the Methodology section.

METHODS

This review followed the PRISMA 2020 reporting guideline (Page et al., 2021); and was designed to make the continuum analytically usable within research on student-AI interaction with LLM-based GenAI in CT development.

Research Design and Database Selection

The review adopted a configurative design, suitable when the evidence base is heterogeneous and the goal is conceptual synthesis rather than pooled effect estimation (Bagdonaitė & Dagienė, 2025; Gough et al., 2017; Otto et al., 2025; Page et al., 2021). Scopus was selected because it offers strong coverage of education and educational-technology journals and supports reproducible metadata filtering through fields such as TITLE-ABS-KEY, document type, language, year, and open-access status (Baas et al., 2020; Pranckutė, 2021; Singh et al., 2021). Single-database searching is acknowledged as a limitation, but the choice aligns with recent AI-in-education reviews (Bagdonaitė & Dagienė, 2025; Huang & Tseng, 2026; Otto et al., 2025).

Search Strategy

A Boolean search string was developed around AI-related technology, CT, and the interactional or pedagogical mechanism. The exact Scopus TITLE-ABS-KEY query was TITLE-ABS-KEY (("artificial intelligence" OR AI OR chatbot* OR "generative AI" OR "large language model*" OR LLM OR ChatGPT) AND ("computational thinking" OR "computational thinking skill*") AND ("problem solving" OR "problem-solving" OR "scaffolding" OR "digital scaffolding" OR "problem-solving process")).

The query combined broad AI and large-language-model terms, canonical CT terminology, and mechanism terms aligned with the continuum. The terms "thinking partner" and "thought partner" were deliberately excluded from the search string because the problem-solver to thinking-partner continuum serves as the analytical lens of this review rather than a selection criterion; retaining these terms could have introduced circularity by privileging studies that already use this vocabulary. Sensitivity checks were conducted to assess the robustness of the search strategy. These checks confirmed that neither removing the potentially circular terms nor adding large-language-model terms altered the final corpus of 14 studies, indicating that corpus selection was robust to these alternative formulations. The query was executed in Scopus on 15 April 2026 at 21:25 (UTC+7) and yielded 212 records.

Eligibility Criteria and Study Selection

Eligibility criteria followed a PICOS-inspired structure adapted for educational reviews (Methley et al., 2014; Page et al., 2021), and study selection proceeded through the PRISMA 2020 stages of identification, screening, and eligibility (Page et al., 2021). Articles were included if they were peer-reviewed journal articles published in English from 2021 onward, openly accessible in full text, and substantively engaged with the use of AI-related technology, with conversational LLM-based GenAI prioritized, to support CT development among students or preservice teachers. The 2021 cut-off was selected to capture both the immediate pre-ChatGPT period, which contains foundational chatbot and intelligent-tutoring-system work that informs the continuum, and the post-November-2022 LLM-based GenAI surge. Conference proceedings, editorials, book chapters, errata, non-English articles, studies with excessively broad keywords signaling tangential engagement, and articles focused exclusively on in-service teachers' professional development without student-level outcomes were excluded.

Identification began with 212 Scopus records retrieved on 15 April 2026. Removing 18 records published before 2021 left 194 records for screening. Title-and-abstract screening then removed 120 non-article items and 1 non-English article, leaving 73 articles for content-relevance screening. Ten records were then excluded as topically tangential, in that their indexed keywords matched the query only through generic terms without substantive engagement with the CT-GenAI intersection, leaving 63 articles for full-text eligibility assessment. Eligibility was assessed in two separable steps: an access-related step (open access and full-text availability) and a content-related step (topic relevance). Full-text assessment excluded 30 paywalled

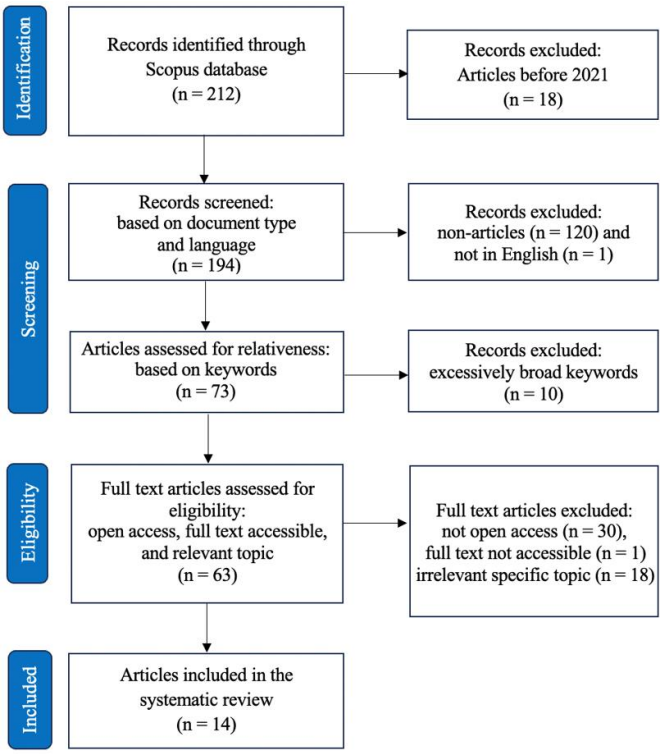


Figure 2. PRISMA 2020 flow diagram for the present SLR (the author's own elaboration)

articles and 1 inaccessible article on access grounds, and 18 off-topic articles on content grounds, yielding a final corpus of 14 peer-reviewed, open-access journal articles published between 2021 and 2026. **Figure 2** reports the exact PRISMA flow.

Data Extraction

A structured extraction template (**Table 2**) was piloted on three articles and applied to the full corpus (Bagdonaitė & Dagienė, 2025; Huang & Tseng, 2026; Otto et al., 2025). The extracted data were then coded using the integrated framework presented in **Table 3**.

Unreported information was coded as "not reported" rather than inferred (Page et al., 2021). Ambiguous continuum placements were resolved through re-reading of methods and results sections and triangulation with the qualitative descriptions provided by the authors.

Data Coding Framework

To translate the theoretical continuum presented in section theoretical framework into an analytically applicable instrument, the review employed an integrated coding framework.

Table 2. Data extraction template aligned with the three RQs

Category	Extracted information	Description and rationale	Linked RQ
Bibliographic	Authors, year, journal, country, sample size, sample characteristics	Provides the descriptive profile of the corpus and supports the demographic synthesis.	Descriptive
Educational context	Educational level, discipline, instructional setting	Captures the cross-level distribution of GenAI–CT studies and supports interpretation of pedagogical safeguards.	**RQ1**, **RQ3**
AI tool and technology class	Specific GenAI tool, classification (LLM-based vs. earlier AI), version where reported	Identifies whether the study examined LLM-based GenAI specifically; supports the substantive scope of the review.	**RQ1**
CT framework and dimensions	Definition of CT used, framework cited, and CT dimensions measured	Supports the harmonization of CT as cognitive process, educational outcome, and interactional practice.	**RQ1**, **RQ2**
Continuum position	Predominant role of AI in the study: problem solver, cognitive tool, or thinking partner	Direct application of the continuum; each article placed at the position that best represents its reported interactions.	**RQ1**
Interaction patterns	Observable student–AI behaviors coded into the four interaction patterns (surface, scaffolded, achiever-linked, dialogic), informed by Otto et al. (2025)	Operationalizes **RQ2** by mapping reported interactions onto the integrated coding framework.	**RQ2**
CT proficiency/ prior experience link	Evidence relating interaction patterns to learners' CT level, prior experience, or AI literacy	Captures the bidirectional relationship between proficiency and interaction quality (Gong et al., 2025; Hu et al., 2025).	**RQ2**
Cognitive benefits	Reported gains in CT or related cognitive constructs	Supports the benefits side of **RQ3**.	**RQ3**
Cognitive risks	Reported risks: cognitive offloading, over-reliance, reduced problem-solving, erosion of independent reasoning	Supports the risks side of **RQ3** and the productive–substitutive offloading distinction.	**RQ3**
Pedagogical safeguards	Recommendations and design principles	Direct input to the safeguards layer of the integrated framework (**Table 3**).	**RQ3**

Table 3. Integrated coding framework linking the continuum to **RQ1**, **RQ2**, and **RQ3**

Coding layer	Continuum position (**RQ1**)	Interaction patterns and categories (**RQ2**)	Cognitive outcomes and risks (**RQ3**)	Pedagogical safeguards (**RQ3**)
Problem solver pole (pole 1)	AI as answer-generator; outsourced cognition; low student agency	Single-query requests; verbatim copying; minimal refinement; predominantly surface patterns	Short-term task completion; risk of substitutive offloading; problem-solving erosion (Liao et al., 2024)	Limit unmediated access; require student justification of AI output; mandatory critique tasks
Cognitive tool mid-point	AI as scaffold; assisted reasoning; moderate student agency	Iterative prompts; explanations and worked examples; stepwise refinement; predominantly scaffolded patterns	Measurable CT gains, especially in algorithmic thinking and creativity (Gong et al., 2025; Hijón-Neira et al., 2024; Liao et al., 2024; Xu et al., 2024)	Generative progressive prompts; AI-generated instructional contexts; structured scaffolding faded over time
Thinking partner pole (pole 2)	AI as dialogic co-thinker; co-construction; high student agency	Debating and summarizing; metacognitive prompts; critique of output; synthesis across turns; predominantly dialogic patterns	Deep CT growth across all dimensions; transfer to new tasks; effects with and effects of the technology (Hsu, 2025; Sánchez Muñoz et al., 2025)	Constructionist prompting; epistemic-vigilance training; reflective dialogue prompts; learner-directed inquiry

Table 3 operationalizes the three continuum positions by linking each position to its corresponding interaction patterns, cognitive outcomes and risks, and pedagogical safeguards, organized so that each coding layer directly serves one of the three RQs.

Analyzing Data

Synthesis followed thematic procedures for configurative reviews (Thomas & Harden, 2008). Coding proceeded in three layers aligned with the RQs. Layer 1 located each study on the continuum (**RQ1**) by reading its reported interaction moves against the operational indicators in **Table 3**; a study was assigned to the problem solver pole when answer-oriented, single-query use with minimal critique predominated, to the cognitive tool mid-point when the AI mainly scaffolded reasoning while the learner retained task ownership, and to the thinking partner pole when sustained, evaluative, dialogic co-construction was evident. Where a study reported different positions for different subgroups or task phases, both positions were recorded rather than averaged, and the split is shown explicitly in **Table 3**. Borderline placements were resolved by re-reading the methods and results and triangulating with the authors' own qualitative descriptions. The first author coded the full corpus and the coding was cross-checked by a second member of the author team, with disagreements resolved through discussion. Layer 2 identified interaction patterns and proficiency links (**RQ2**); layer 3 clustered benefits, risks, and safeguards (**RQ3**) interpreted through the productive-substitutive offloading distinction (Gong et al., 2025; Hijón-Neira et al., 2024; Hsu, 2025; Hu et al., 2025; Otto et al., 2025; Sánchez Muñoz et al., 2025; Wu & Yang, 2022). The design is constrained by Scopus-only searching, the exclusion of 30 paywalled articles, which may introduce selection bias against paywalled computer-science education venues such as ACM and IEEE, and a search closure date of 15 April 2026. These constraints are revisited in the limitations.

RESULTS

Results are organized by the three RQs. **Table 4** provides the master synthesis of the corpus.

Table 4. Master synthesis of the 14 included studies mapped to **RQ1**, **RQ2**, and **RQ3**

Study (role)	Method and sample	**RQ1**: Continuum position	**RQ2**: Interaction pattern	**RQ3**: Key B/R/S
#1 Gong et al. (2025) [Anchor]	Quant. learning-analytics quasi-exp.; 44 undergrads (China)	Mid → Pole 2 (split by CT level)	Generative progressive prompting; surface (guided/exploratory) vs. in-depth (debating, summarizing)	B: CT and all 5 sub-dim ↑; algorithmic thinking central; R: Lower-CT students stuck in surface dialogue; S: Generative progressive prompts
#2 Liao et al. (2024) [Anchor]	Mixed-method quasi-exp. (7 weeks); 44 sophomores (China)	Mid (cognitive tool)	IPSSC scaffolding via solution assessment, code assessment, free interaction	B: CT, creativity, algorithmic thinking ↑; R: Problem-solving ↓ (paradox); S: Structured scaffolding within bounded tasks
#3 Yunianto et al. (2024) [Anchor]	Educational design research (cycle 4); 5 STEAM scholars (Austria)	Mid → Pole 1 under low prior knowledge	Prompt-quality dependence; algorithmic and debugging support; affective prompts unused	B: Algorithmic and debugging support; R: Strong sensitivity to prompting quality and prior knowledge; S: Pre-requisite domain knowledge
#4 Hsu (2025) [Anchor]	Conceptual/theoretical synthesis (no empirical n)	Pole 2 (theoretical target)	Five-principle constructionist prompting: meaningful,	B: Maps CT elements to prompting practices; S: Constructionist prompting

Table 4 (Continued).

Study (role)	Method and sample	**RQ1**: Continuum position	**RQ2**: Interaction pattern	**RQ3**: Key B/R/S
			iterative, social, metacognitive, learner-directed	B: framework; teacher-education workshop design
#5 Otto et al. (2025) [Anchor]	PRISMA SLR; 50 STEM studies (2023–2024)	Spans full continuum	Five categories: tutoring, co-creating (dominant), processing, coaching, simulating	B: CT, critical thinking, problem-solving supported; R: Algorithmic bias, reliability, pedagogical demand; S: Pedagogical orchestration
#6 Sánchez Muñoz et al. (2025) [Anchor]	Qualitative action-research (14 weeks); 24 pre-university (Mexico)	Mid → Pole 2 (progression over weeks)	Progression from content-generation to metacognitive and dialogic use; prompt formulation	B: Prompt formulation, ideation, computational fluency ↑; R: Epistemic vigilance does not emerge automatically; S: Critical-reflection scaffolds
#7 Wu and Yang (2022) [Anchor]	One-group pre-/post-test (6h); 22 primary students (Taiwan)	Mid (with teacher) vs. pole 1/2 (without)	Teacher-mediated help-seeking vs. independent exploration	B: Cognitive understanding ↑; R: Teacher support increases dependence; S: Graduated release of teacher support
#8 Hu et al. (2025) [Anchor]	Cross-sectional survey + SEM; 1,466 univ. students (China, 4 'new' disciplines)	Spans continuum (AI-literacy proxy)	AI-literacy dimensions associated with CT; small but significant effects; discipline-specific	B: AI literacy positively associated with CT; R: Effect sizes small; non-linear use patterns; S: AI-literacy training
#9 Bagdonaitė and Dagienė (2025) [Anchor]	PRISMA SLR; 94 primary-education studies (2020–2025)	Spans continuum	Adaptive feedback, inquiry scaffolding, CT-supportive analytics	B: Reading, problem-solving, engagement, CT supported; R: TPACK gaps, infrastructural inequity; S: Pedagogically grounded design > technical sophistication
#10 Hijón-Neira et al. (2024) [Anchor]	Experiment (3-week); 122 preservice ECE teachers (Spain)	Mid (cognitive tool)	AI-generated contexts for robotics tasks; affective engagement	B: Problem-solving and algorithmic thinking ↑; technology acceptance ↑; S: AI-generated instructional contexts
#11 Omeh et al. (2025) [Empirical]	Quasi-exp. with retention; 120 univ. programming students (Nigeria)	Mid (cognitive tool)	CBL/PBL pedagogy with GitHub Copilot	B: CT and reflective thinking ↑ (large effect, d = 0.542); achievement and retention ↑; S: Innovative pedagogy with Copilot embedded
#12 Xu et al. (2024) [Empirical]	Mixed-method quasi-exp.; 77 primary students (China)	Mid → Pole 1 (low-achievers) vs. pole 2 (high-achievers)	Selective dialogue and confirmation loops (high) vs. free dialogue and direct clue-to-answer (low)	B: CT, problem-solving, flow ↑; R: Achiever-level interaction divergence; S: Achiever-adaptive design (game-based)
#13 Huang and Tseng (2026) [Empirical]	PRISMA SLR (thematic); programming education (Jan 2023-May 2025)	Mid (most common in literature)	Skill-level moderates benefit; foundational vs. complex tasks differ	B: Knowledge, code quality, completion speed ↑; R: Depth, diversity, debug quality ↓; S: Task-difficulty calibration
#14 Ameen et al. (2024) [Empirical]	Experimental (pre/post); 146 students (Iraq; Chem + CS)	Mid (cognitive tool)	ChatGPT-supported teaching across two disciplines	B: Achievement and CT ↑ (large r = 0.54); S: Cross-disciplinary applicability

Note. B: Benefits; R: Risks; S: Safeguards; & Continuum positions reflect the predominant or modal role of LLM-based GenAI as reported in each study, spanning studies are reviews whose included corpus covered multiple positions

Figure 3 visualizes the overall logic that connects the three sets of findings.

Overview of Included Studies

The corpus spans 2021-2026 and covers primary, secondary, pre-university, undergraduate, graduate or adult, preservice teacher, and cross-disciplinary university contexts (Ameen et al., 2024; Gong et al., 2025; Hijón-Neira et al., 2024; Hu et al., 2025; Liao et al., 2024; Sánchez Muñoz et al., 2025; Wu & Yang, 2022; Xu et al., 2024; Yunianto et al., 2024). The methods used include quasi-experiments, design research, an SEM survey, action research, conceptual synthesis, and systematic reviews (Bagdonaitė & Dagienė, 2025; Huang & Tseng, 2026; Otto et al., 2025).

RQ1: Continuum Positioning of LLM-Based GenAI in Computational Thinking Learning

Across the 14 studies, LLM-based GenAI was positioned along the full problem-solver-to-thinking-partner continuum, although most studies clustered around the cognitive tool mid-point. In this position, GenAI was used through structured scaffolding, embedded chatbots, or pedagogically designed activities that supported explanation, feedback, debugging, and refinement rather than simply producing final answers (Ameen et al., 2024; Hijón-Neira et al., 2024; Hu et al., 2025; Huang & Tseng, 2026; Liao et al., 2024; Omeh et al., 2025; Otto et al., 2025). Positioning closer to the thinking partner pole took two analytically distinct forms. A small number of studies reported genuine change over time, in which students developed more dialogic and metacognitive engagement across a task sequence. More commonly, studies reported between-group differences observed cross-sectionally, in which higher-proficiency students were more likely to critique, verify, and extend AI responses than lower-proficiency peers; these differences describe how groups differ at a single point rather than movement of the same learners over time (Gong et al., 2025; Hsu, 2025; Sánchez Muñoz et al., 2025; Xu et al., 2024). Positioning closer to the problem solver pole appeared, as a between-group difference, when prior knowledge, achievement level, or instructional

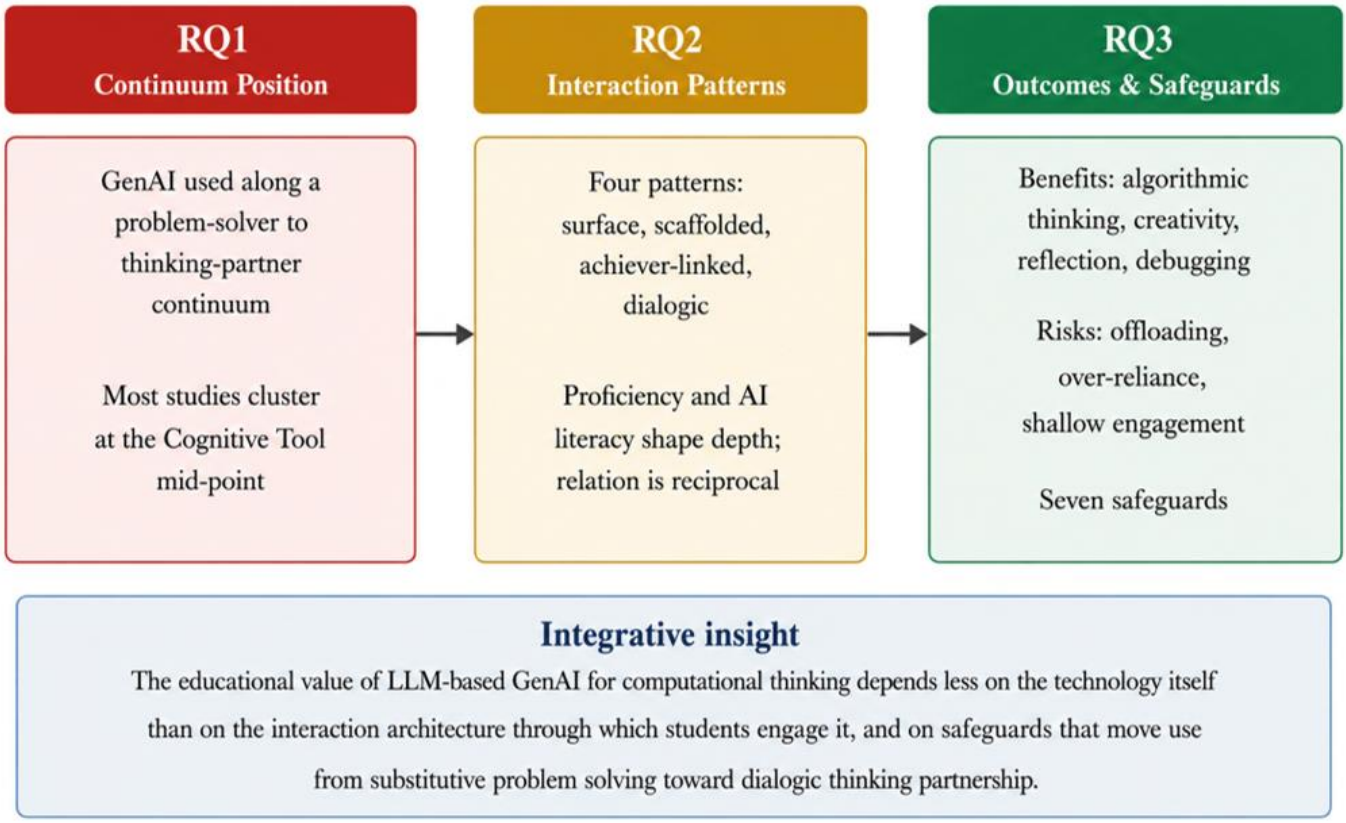


Figure 3. Synthesis of findings across **RQ1**, **RQ2**, and **RQ3** (the author's own elaboration)

support was weak, as observed among low-achievers, learners with limited GeoGebra command knowledge, and primary students working without teacher support (Wu & Yang, 2022; Xu et al., 2024; Yunianto et al., 2024).

Two patterns follow from this corpus map. First, the continuum describes interaction quality rather than tool identity: the same LLM-based tool, most often ChatGPT, could occupy the problem solver, Cognitive Tool, or thinking partner position depending on task design, student profile, and pedagogical support (Bagdonaitė & Dagienė, 2025; Hsu, 2025; Otto et al., 2025). Second, continuum positioning varied with conditions rather than being fixed: interaction was positioned closer to the thinking partner pole when scaffolds encouraged reflection and dialogue but drifted toward the problem solver pole when those supports were absent (Gong et al., 2025; Sánchez Muñoz et al., 2025; Wu & Yang, 2022).

RQ2: Student-AI Interaction Patterns and Their Linkage to CT Proficiency

Across the corpus, four broad classes of student-AI interaction patterns were identified: surface, scaffolded, achiever-linked, and dialogic. Surface patterns involved single-query requests, verbatim copying, unstructured free dialogue, and direct clue-to-answer transitions. These patterns appeared more frequently among lower-CT or low-achieving students, who tended to use GenAI for immediate support rather than sustained reasoning (Gong et al., 2025; Wu & Yang, 2022; Xu et al., 2024). Scaffolded patterns included iterative prompting, requests for explanations, worked examples, debugging, and stepwise refinement; these were common in cognitive tool studies in which GenAI supported reasoning without replacing it (Ameen et al., 2024; Hijón-Neira et al., 2024; Liao et al., 2024; Omeh et al., 2025; Otto et al., 2025).

Achiever-linked patterns were most visible in which high-achieving students used selective dialogue and confirmation loops to verify and refine their reasoning, whereas low-achieving students tended to rely on immediate-answer pathways (Xu et al., 2024). Interaction quality was also shaped by factors such as CT proficiency, prior experience, and AI literacy (Gong et al., 2025; Hu et al., 2025). Dialogic patterns, including debating AI output, summarizing, critique, metacognitive prompting, and synthesis across turns, were observed mainly among higher-CT students and within structured progressions toward the thinking partner pole (Gong et al., 2025; Hsu, 2025; Sánchez Muñoz et al., 2025). The relationship between CT proficiency and interaction quality is reciprocal: higher CT and AI literacy enable deeper dialogue, while scaffolded dialogue can further strengthen CT growth (Gong et al., 2025; Hu et al., 2025; Liao et al., 2024; Omeh et al., 2025; Sánchez Muñoz et al., 2025). Without scaffolds, lower-CT students risk remaining at the problem solver pole, an outcome with significant equity implications (Bagdonaitė & Dagienė, 2025; Hu et al., 2025).

Table 5. Pedagogical safeguards synthesized from the corpus, mapped to the continuum

Safeguard	Operational description	Primary target position	Key sources
Constructionist prompting	Train students to craft, test, revise, and reflect on prompts as meaningful, iterative, social, metacognitive, and learner-directed practices.	Thinking partner pole	Hsu (2025)
Generative progressive prompts	Use staged prompts that begin with guided exploration and progressively elicit debating, summarizing, and metacognitive moves.	Cognitive tool → thinking partner	Gong et al. (2025)
Critical-reflection scaffolds	Embed structured reflection that requires evaluating AI output, justifying acceptance or revision, and articulating reasoning.	Thinking partner pole	Sánchez Muñoz et al. (2025)
AI-literacy training	Build the foundational AI literacy and metacognitive base required for deeper student–AI dialogue, especially in non-STEM programs.	Problem solver → cognitive tool	Hu et al. (2025)
Graduated release of teacher support	Provide initial teacher mediation, then withdraw scaffolding incrementally so students learn to engage AI independently without dependence.	Cognitive tool → thinking partner	Wu and Yang (2022)
Achiever-adaptive design	Differentiate task structures and prompt scaffolds for higher- and lower-achievers, recognizing that the same tool will be used very differently.	Problem solver → cognitive tool	Gong et al. (2025), Xu et al. (2024)
Task-difficulty and prior-knowledge calibration	Match task complexity to learner readiness; ensure adequate domain knowledge before relying on GenAI for complex tasks.	Cognitive tool mid-point	Huang and Tseng (2026), Yunianto et al. (2024)

RQ3: Cognitive Benefits, Risks, and Pedagogical Safeguards

Cognitive benefits clustered around algorithmic thinking, creativity, reflective and critical thinking, engagement, motivation, and flow. Algorithmic thinking was the most consistent gain, while reflection-related gains were strongest when structured reflection was embedded (Ameen et al., 2024; Gong et al., 2025; Hijón-Neira et al., 2024; Liao et al., 2024; Omeh et al., 2025; Sánchez Muñoz et al., 2025; Xu et al., 2024). Cognitive risks included problem-solving erosion, dependence on scaffolding, reduced depth and diversity of solutions, weaker debugging quality, surface engagement, prompting fragility under weak prior knowledge, and broader concerns of algorithmic bias, reliability, and infrastructural inequity (Bagdonaitė & Dagienė, 2025; Huang & Tseng, 2026; Liao et al., 2024; Otto et al., 2025; Wu & Yang, 2022; Yunianto et al., 2024). Taken collectively, these risks instantiate substitutive cognitive offloading: the transfer of cognitive responsibility from student to AI, rather than the productive offloading that frees working-memory resources while preserving student agency.

Seven safeguard families emerged from the corpus: constructionist prompting, generative progressive prompts, critical-reflection scaffolds, AI-literacy training, graduated release of teacher support, achiever-adaptive design, and task-difficulty and prior-knowledge calibration. **Table 5** maps each safeguard to the continuum position at which it predominantly operates.

Figure 3 visualizes the cohering logic of the three RQs: continuum positioning (**RQ1**) shapes the interaction patterns that emerge (**RQ2**), which in turn drive the cognitive benefits, risks, and pedagogical safeguards reported across the corpus (**RQ3**).

DISCUSSION

The synthesis supports a central argument: the educational value of LLM-based GenAI for CT development depends less on the tool itself than on the quality of student-AI interaction. Across the 14 included studies, GenAI was not consistently positioned as either a problem-solving tool or a thinking partner. Instead, its role shifted along the problem-solver-to-thinking-partner continuum, depending on task design, student profile, prompt strategy, AI literacy, and pedagogical scaffolding. This finding confirms that the continuum describes interaction quality rather than tool identity. The same LLM-based GenAI system, most often ChatGPT, may occupy the problem solver, Cognitive Tool, or thinking partner position in CT learning (Bagdonaitė & Dagienė, 2025; Gong et al., 2025; Hsu, 2025; Otto et al., 2025).

The continuum also clarifies several empirical tensions in the literature. Studies have reported CT gains alongside decreased problem-solving when students used a ChatGPT-powered scaffold, suggesting that the intervention generated effects with the technology but did not fully produce effects of the technology because students may not have internalized the problem-solving processes supported by AI (Liao et al., 2024; Salomon et al., 1991). Similarly, teacher-supported AI activities have been shown to improve immediate learning outcomes while also increasing dependence, indicating that scaffolding may support learning while still anchoring

students at the cognitive tool mid-point when epistemic agency is not gradually transferred back to learners (Wu & Yang, 2022).

The findings further show that student proficiency and AI literacy shape the depth of student-AI interaction. Higher-CT students tend to engage in selective dialogue, verification, critique, synthesis, and metacognitive prompting. Lower-CT students are more likely to rely on direct answers, surface dialogue, and immediate clue-to-answer pathways. This pattern suggests a reciprocal relationship: stronger CT and AI literacy enable deeper interaction, while well-scaffolded interaction can further develop CT (Gong et al., 2025; Hu et al., 2025; Xu et al., 2024). Without appropriate safeguards, LLM-based GenAI may widen rather than reduce learning disparities, because students with weaker prior knowledge are more vulnerable to substitutive cognitive offloading (Bagdonaitė & Dagienė, 2025; Hu et al., 2025).

These results shift the pedagogical question from tool adoption to interaction-architecture design. Effective use of LLM-based GenAI requires task structures, prompt scaffolds, reflection routines, and teacher orchestration that move students from answer-seeking toward critique, revision, and synthesis. The seven safeguards identified in this review provide practical directions for this shift: constructionist prompting, generative progressive prompts, critical-reflection scaffolds, AI-literacy training, graduated teacher support, achiever-adaptive design, and task-difficulty calibration. These safeguards are not isolated techniques, but complementary design principles for preserving epistemic agency while using GenAI to support CT development (Gong et al., 2025; Hsu, 2025; Sánchez Muñoz et al., 2025; Wu & Yang, 2022).

The review should also be interpreted in light of two boundaries. The first concerns database scope. Relying on Scopus alone, although justified by its broad and curated coverage of education and educational-technology journals and its reproducible metadata filtering (Baas et al., 2020; Pranckutė, 2021; Singh et al., 2021), may under-represent venues indexed elsewhere. Incorporating Web of Science or ERIC could have added paywalled computer-science outlets such as ACM and IEEE proceedings and some education-specific records; this would most plausibly broaden the corpus and reinforce, rather than overturn, the observed patterns, so the single-database design is best read as a boundary on generalizability rather than a threat to the internal validity of the synthesis. The second boundary concerns the evidence base itself. Many included studies used short-duration interventions, pre/post designs, or limited interaction-pattern analysis. Framed through the distinction between effects with and effects of technology (Salomon et al., 1991), such designs can evidence effects with GenAI, namely enhanced performance while the tool is in use, but cannot establish effects of GenAI, namely durable cognitive change that persists after the tool is withdrawn. Longitudinal and delayed-post-test designs are therefore still needed to determine whether thinking partner interactions produce durable effects of the technology. Future research should examine how specific safeguards influence student-AI interaction over time and should expand beyond ChatGPT-centric designs to include multimodal and agentic GenAI systems.

Effective CT instruction should move beyond tool adoption toward interaction-architecture design that includes constructionist prompting, progressive prompts, critical-reflection scaffolds, AI-literacy training, graduated teacher support, adaptive task design, and careful calibration of task difficulty. In practice, LLM-based GenAI should be used to support explanation, critique, revision, and synthesis rather than to replace student reasoning.

CONCLUSIONS

This SLR shows that LLM-based GenAI is neither inherently a problem-solving tool nor a thinking partner; its educational value depends on how students engage it during CT learning. Across the corpus, GenAI use spans a continuum from the problem solver pole, through a cognitive tool mid-point, to the thinking partner pole. Most studies locate it at the mid-point, where it supports explanation, scaffolding, feedback, debugging, and refinement; students drift toward the problem solver pole when seeking direct answers and toward the thinking partner pole when questioning, critiquing, and extending their own reasoning.

Four interaction patterns recur: surface, scaffolded, achiever-linked, and dialogic. The relationship between engagement and competence is reciprocal: stronger CT and AI literacy enable deeper, more dialogic interaction, while well-designed interaction further develops CT. Lower-proficiency learners more often rely on answer-seeking exchanges associated with substitutive offloading, over-reliance, shallow engagement, and unequal outcomes. These risks stem not from the tool but from interaction designs that let GenAI replace rather than strengthen reasoning. The central pedagogical challenge is therefore not whether students use GenAI, but how they are guided to use it. With deliberate interaction-architecture design and embedded safeguards, GenAI is more likely to cultivate CT than to substitute for it.

Author contributions: MM: conceptualization, data curation, formal analysis, methodology, visualization, writing – original draft, writing – review & editing; **TH:** conceptualization, methodology, supervision, validation, writing – review & editing; **YSK:** methodology, supervision, validation, writing – review & editing; **NN:** supervision, validation, writing – review & editing; **TGM:** data curation, investigation, writing – review & editing. All authors agreed with the results and conclusions.

Funding: This study was supported by Indonesian Education Scholarship, Center for Higher Education Funding and

Assessment, Kemdiktisaintek and Indonesian Endowment Fund for Education.

Acknowledgments: The authors would like to thank the colleagues at the department of mathematics education, Indonesia University of Education, for their valuable feedback during the preparation of this review, and all individuals and institutions whose support and contributions made this research possible.

Ethical statement: This study is a SLR based exclusively on previously published sources and did not involve direct participation of human subjects.

AI statement: The authors declare that no GenAI was used in the creation of this manuscript.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Ameen, L. T., Yousif, M. R., Alnoori, N. A. J., & Majeed, B. H. (2024). The impact of artificial intelligence on computational thinking in education at university. *International Journal of Engineering Pedagogy, 14*(5), 192-203. https://doi.org/10.3991/ijep.v14i5.49995

Arum, R., Calderon Leon, M., Li, X., & Lopes, J. (2025). ChatGPT early adoption in higher education: Variation in student usage, instructional support, and educational equity. *AERA Open, 11*. https://doi.org/10.1177/23328584251331956

Baas, J., Schotten, M., Plume, A. M., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies, 1*(1), 377-386. https://doi.org/10.1162/qss_a_00019

Bagdonaitė, J., & Dagienė, V. (2025). Artificial intelligence in primary education: A systematic literature review 2020-2025. *Informatics in Education, 24*(4), 697-736. https://doi.org/10.15388/infedu.2025.24

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *PNAS, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 Annual Meeting of the American Educational Research Association* (pp. 1-25).

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7-19. https://doi.org/10.1093/analys/58.1.7

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), Article 6. https://doi.org/10.3390/soc15010006

Gong, X., Xu, W., & Qiao, A. (2025). Exploring undergraduates' computational thinking and human-computer interaction patterns in generative progressive prompt-assisted programming learning. *International Journal of Educational Technology in Higher Education, 22*, Article 51. https://doi.org/10.1186/s41239-025-00552-y

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd ed.). SAGE. https://doi.org/10.4135/9781036234942.n1

Grassini, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology, 16*. https://doi.org/10.3389/fpsyg.2025.1550621

Hijón-Neira, R., Pizarro, C., Borrás-Gené, O., & Cavero, S. (2024). AI-generated context for teaching robotics to improve computational thinking in early childhood education. *Education Sciences, 14*(12), Article 1401. https://doi.org/10.3390/educsci14121401

Hsu, H.-P. (2025). From programming to prompting: Developing computational thinking through large language model-based generative artificial intelligence. *TechTrends, 69*, 485-506. https://doi.org/10.1007/s11528-025-01052-6

Hu, Z., He, H., Zhang, C., & Guan, Y. (2025). Development and influencing factors of artificial intelligence literacy and computational thinking in Chinese university students. *Scientific Reports, 15*, Article 42708. https://doi.org/10.1038/s41598-025-26888-z

Huang, T.-C., & Tseng, H.-P. (2026). Learning, behavior, and pedagogy: A systematic review of generative AI use in programming education. *International Journal of Information and Education Technology, 16*(1), 102-113. https://doi.org/10.18178/ijiet.2026.16.1.2487

Ilic, U., Haseski, H. İ., & Tugtekin, U. (2018). Publication trends over 10 years of computational thinking research. *Contemporary Educational Technology, 9*(2), 131-153. https://doi.org/10.30935/cet.414798

Kıyıcı, G., & Yamak, H. (2026). Effects of science education on computational thinking: A meta-analysis. *European Journal of Science and Mathematics Education, 14*(4), 525-546. https://doi.org/10.30935/scimath/18904

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

Liao, J., Zhong, L., Zhe, L., Xu, H., Liu, M., & Xie, T. (2024). Scaffolding computational thinking with ChatGPT. *IEEE Transactions on Learning Technologies, 17*, 1628-1642. https://doi.org/10.1109/TLT.2024.3392896

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research, 14*, Article 579. https://doi.org/10.1186/s12913-014-0579-0

Omeh, C. B., Olelewe, C. J., & Ohanu, I. B. (2025). Impact of artificial intelligence technology on students' computational and reflective thinking in a computer programming course. *Computer Applications in Engineering Education, 33*, Article e70052. https://doi.org/10.1002/cae.70052

Otto, S., Lavi, R., & Bertel, L. B. (2025). Human-GenAI interaction for active learning in STEM education: State-of-the-art and future directions. *Computers & Education, 239*, Article 105444. https://doi.org/10.1016/j.compedu.2025.105444

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Pea, R. D. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational Psychologist, 20*(4), 167-182. https://doi.org/10.1207/s15326985ep2004_2

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). Cambridge University Press.

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications, 9*(1), Article 12. https://doi.org/10.3390/publications9010012

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Rohaeti, E., & Huda, N. (2025). Assessing computational thinking skills of science and mathematics upper-secondary school students. *European Journal of Science and Mathematics Education, 13*(4), 289-303. https://doi.org/10.30935/scimath/17248

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher, 20*(3), 2-9. https://doi.org/10.3102/0013189X020003002

Sánchez Muñoz, J. A., Flores-Eraña, G., Silva-Campos, J. M., Chavira-Quintero, R., & Olais-Govea, J. M. (2025). GenAI as a cognitive mediator: A critical-constructivist inquiry into computational thinking in pre-university education. *Frontiers in Education, 10*. https://doi.org/10.3389/feduc.2025.1597249

Sendarrubias-Hipólito, M., Agreda-Montoro, M., Rodríguez-Moreno, J., & Ortiz-Colón, A. M. (2026). The integration of artificial intelligence, robotics, and computational thinking in education: Overview and trends. *Contemporary Educational Technology, 18*(2), Article ep660. https://doi.org/10.30935/cedtech/18600

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142-158. https://doi.org/10.1016/j.edurev.2017.09.003

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics, 126*(6), 5113-5142. https://doi.org/10.1007/s11192-021-03948-5

Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments, 32*(9), 5142-5155. https://doi.org/10.1080/10494820.2023.2209881

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

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A, 366*(1881), 3717-3725. https://doi.org/10.1098/rsta.2008.0118

Wu, S.-Y., & Yang, K.-K. (2022). The effectiveness of teacher support for students' learning of artificial intelligence popular science activities. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.868623

Xu, Y., Zhu, J., Wang, M., Qian, F., Yang, Y., & Zhang, J. (2024). The impact of a digital game-based AI chatbot on students' academic performance, higher-order thinking, and behavioral patterns in an information technology curriculum. *Applied Sciences, 14*(15), Article 6418. https://doi.org/10.3390/app14156418

Yunianto, W., Lavicza, Z., Kastner-Hauler, O., & Houghton, T. (2024). Investigating the use of ChatGPT to solve a GeoGebra-based mathematics + computational thinking task in a geometry topic. *Journal on Mathematics Education, 15*(3), 1027-1052. https://doi.org/10.22342/jme.v15i3.pp1027-1052

Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education, 21*, Article 21. https://doi.org/10.1186/s41239-024-00453-6