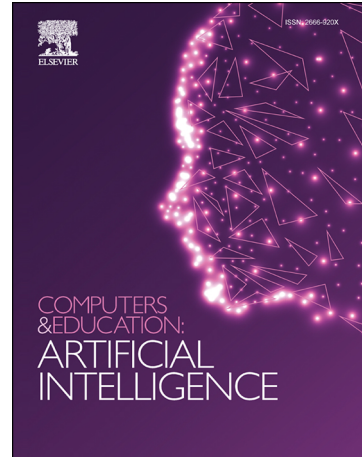


## Journal Pre-proof

Generative AI-enhanced learning experiences for computational thinking: A systematic scoping review and design guidelines

Abdessalam Ouazaki, Antonette Shibani, Simon Knight, Adrian Holzer

PII: S2666-920X(26)00070-6  
DOI: <https://doi.org/10.1016/j.caeai.2026.100608>  
Reference: CAEAI 100608



To appear in: *Computers and Education: Artificial Intelligence*

Received Date: 13 August 2025  
Revised Date: 28 April 2026  
Accepted Date: 4 May 2026

Please cite this article as: Ouazaki, A., Shibani, A., Knight, S., & Holzer, A. (2026). Generative AI-enhanced learning experiences for computational thinking: A systematic scoping review and design guidelines. *Computers and Education: Artificial Intelligence*. doi: <https://doi.org/10.1016/j.caeai.2026.100608>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see:

<https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>.

Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Generative AI-Enhanced Learning Experiences for Computational Thinking: A Systematic Scoping Review and Design Guidelines

Abdessalam Ouazki<sup>a,\*</sup>, Antonette Shibani<sup>b</sup>, Simon Knight<sup>b</sup>, Adrian Holzer<sup>a</sup>

<sup>a</sup>*University of Neuchâtel, Neuchâtel, Switzerland*

<sup>b</sup>*University of Technology Sydney (UTS), Sydney, Australia*

---

---

## Acknowledgments

This research was partially supported by the Swiss National Science Foundation (SNSF) under project number IZSEZO\_225217, and by Swissuniversities PgB project *STAGE: STEM Teaching in the Age of Generative AI*.

---

\*Corresponding author.

*Email addresses:* [abdessalam.ouazki@unine.ch](mailto:abdessalam.ouazki@unine.ch) (Abdessalam Ouazki),  
[Antonette.Shibani@uts.edu.au](mailto:Antonette.Shibani@uts.edu.au) (Antonette Shibani), [Simon.Knight@uts.edu.au](mailto:Simon.Knight@uts.edu.au)  
(Simon Knight), [adrian.holzer@unine.ch](mailto:adrian.holzer@unine.ch) (Adrian Holzer)

# Generative AI-Enhanced Learning Experiences for Computational Thinking: A Systematic Scoping Review and Design Guidelines

---

## Abstract

Computational thinking (CT) skills are crucial to teach. The rise of Generative AI could potentially revolutionize the teaching of these skills by changing learning experiences and outcomes. Educators are rapidly integrating these tools into their educational activities; however, the best practices for doing so remain unclear. Existing reviews on GenAI are either too broad in regards to studying specific tools such as ChatGPT for learning, or too limited in terms of studying computational thinking with a focus on programming. As such, this paper aims to summarize existing knowledge and extract best practices and avenues for future research and practice around GenAI and computational thinking skills. Our contributions are two-fold: 1) Insights from a systematic scoping review of studies examining the use of GenAI to support the teaching of CT skills in diverse contexts and their reported learning outcomes, 2) Design guidelines to inform effective use of GenAI in CT education for pedagogic practice. Our results reveal a young but rapidly growing research field, with most interventions focusing on undergraduate students and basic programming tasks, often using off-the-shelf tools with limited integration. GenAI is typically used as a coder, tutor, debugger, or ideator, with mixed effects on learning outcomes. A key challenge is the ten-

sion between overreliance by beginners, who may offload thinking to GenAI, and under-utilization by advanced learners in complex projects. We derive seven guidelines to provide actionable recommendations from the findings to effectively integrate GenAI for CT while minimizing associated risks and guiding responsible AI use in education, while also suggesting directions to inform the design of future systems and CT pedagogies.

---

## 1. Introduction

A little more than two years ago, the conversational agent revolution seemed a distant hype. However, with the release of Generative Artificial Intelligence (GenAI)-enabled agents such as ChatGPT, released in late 2022, the world was provided with a proof-of-concept showing the potential (and risks) of future conversational systems. Conversational agents are increasingly integrated in various domains from healthcare to education, utilizing the human-like capabilities of GenAI that ride the latest “AI wave” (Schöbel et al., 2024). As such, this type of interaction shows promise and deserves to be investigated in more detail. The literature on conversational agents itself is still nascent, and several important challenges remain. For instance, it is important to measure and assess these agents more *holistically* to determine their actual benefits to users (Følstad et al., 2021). This evaluation should take into consideration aspects such as usefulness, efficiency, and process support (Følstad et al., 2021; Rapp et al., 2021), as well as psychological outcomes such as motivation, anxiety, performance, and perception of using these systems (Okonkwo and Ade-Ibijola, 2021). Despite the growing number of studies on user experience with these technologies, there is a lack of

standardized definitions, metrics, and validated scales for evaluating key aspects of conversational agents, and a lack of commonly applied approaches to evaluation (Følstad et al., 2021).

### *1.1. Conversational agents as learning technologies*

In the educational context, a recent review of the literature shows an exponentially growing interest in the use of conversational agents since 2016 (Okonkwo and Ade-Ibijola, 2021). These agents can assist with content integration, providing quick access to educational information, increasing motivation and engagement, allowing multiple users to access the system at the same time, and providing immediate assistance (Okonkwo and Ade-Ibijola, 2021). They have been successfully implemented in various educational contexts, aiding in time-saving (Ranoliya et al., 2017), maximizing student learning abilities, and increasing student engagement (Clarizia et al., 2018). Conversational agents have also been designed and deployed to function as virtual assistants helping with administrative and academic tasks (Sinha et al., 2020) or answering frequently asked questions (FAQs) to personalize and enrich the user experience (Peyton and Unnikrishnan, 2023; Ranoliya et al., 2017; Sethi and Farhana, 2020).

The rise of GenAI has elevated the usefulness and value of such agents, as it enables the development of personalized learning experiences that cater to individual student needs, resulting in enhanced learning outcomes (Su and Yang, 2023; Ouazaki et al., 2024). It also boosts efficiency and productivity by facilitating task automation and encouraging human-AI collaboration (Sun et al., 2024; Peña-Fernández et al., 2023). Recent reviews, including a systematic review and meta-analysis by Deng et al. (Deng et al.,

2024), suggest that ChatGPT interventions, primarily in university classrooms, may support academic performance, motivation, and higher-order thinking, while potentially reducing cognitive load. However, the strength of such early evidence remains contested. For instance, Weidlich et al. (Weidlich and Gašević, 2025) raised concerns about the methodological limitations of studies included in such meta-analyses, noting issues such as the absence of random assignment, power analyses, and pre-tests, as well as broader conceptual ambiguities in study designs. A recent study also showed that while tools such as ChatGPT can improve short-term task performance, they may not boost knowledge gain and transfer, and may trigger metacognitive laziness (Fan et al., 2025), recommending GenAI integration in teaching and learning be done with caution.

### *1.2. GenAI and computational thinking abilities*

One aspect of education that is particularly impacted by the GenAI revolution is the teaching and learning of computational thinking (CT) skills. CT has been touted as a fundamental skill to be added to one's analytical ability as it involves "*solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science*" (Wing and M, 2006). With many definitions and identified elements of CT, such as abstraction, generalization, decomposition, algorithmic thinking, and debugging (Angeli et al., 2016; Hsu, 2025), the need to incorporate CT in curricula is highlighted across educational contexts for learners of different age groups. The relevance of this particular topic stems from the fact that with the massive digital transformation of our lives, teaching CT abilities has become crucial for all students, not just those who wish to become software

engineers or computer scientists (Wing and Jeannette, 2011).

With GenAI, conversational agents have significant potential for understanding, reviewing, revising, and creating code across languages (Denny et al., 2024), and are thought to become major game changers for programming. These agents are now used by many developers, fundamentally changing their development processes and workflows (Khemka and Houck, 2024; Stack Overflow, 2023). They have the potential to shift how we conceptualize programming, towards humans as “relegated to, at best, a supervisory role”, overseeing AI generated software (Welsh and Matt, 2022). As such, CT skills, i.e., the ability to structure a real-life problem in a way that allows for its automation, are not obsolete in themselves. In this context, existing literature has investigated how GenAI can be used in different computational thinking learning scenarios and courses, such as data-oriented programming, Python for data analysis, and software testing (Yilmaz and Yilmaz, 2023; Ouazaki et al., 2023; Gottipati et al., 2023) However, learner interactions with technologies might need revising in this context, including towards learning experiences that integrate conversational agents, and learning objectives that focus on quality control of AI-generated code rather than low-level code creation (Denny et al., 2024).

While computational thinking has never been solely about programming, its emphasis on higher-order problem-solving skills becomes even more critical in the context of GenAI integration, both to fully leverage this technology and to navigate its ethical implications. Because CT underpins how people evaluate evidence, reason with data, and critically judge outputs, exploring how GenAI transforms CT education is timely and consequential. GenAI

may shift learners from constructing reasoning to curating and verifying it, as it can reorient critical thinking toward oversight rather than primary content generation Lee et al. (2025). However, it is important to understand how this shift influences learners' longer-term reasoning capabilities, particularly if they outsource these core processes completely or have fewer opportunities to develop them in the first place. Developing CT (independent of GenAI support) may therefore offer a “last-mile” pathway to societal impact, with the potential to support more informed decision-making beyond the university and widen access to digital reasoning skills that shape participation in technology-rich futures.

### *1.3. Study contribution*

This study aims to support educators and researchers better understand the landscape of GenAI in the particular context of teaching CT skills. More specifically, this literature review addresses the following research questions:

- RQ1: In which contexts are GenAI solutions for teaching CT skills deployed?
- RQ2: What types of GenAI solutions for teaching CT skills have been developed?
- RQ3: How mature are GenAI solutions for teaching CT skills in terms of their evaluation?
- RQ4: What concerns and limitations are reported about the use of GenAI solutions for teaching CT skills, and what recommendations have been proposed for their optimal use?

The study then builds on the findings to provide actionable guidelines to help educators in their practice and to help academics to set their research agendas, as well as to sketch directions for how CT education may need to evolve in response to the growing capabilities of GenAI.

Even though several recent literature reviews have examined the broader educational impact of GenAI, none has yet provided a specific investigation into the design of GenAI enhanced learning experiences for CT skills. For instance, Deng et al.'s meta-analysis focused on ChatGPT's effects on general student learning outcomes (Deng et al., 2024), Batista et al.'s synthesis reviewed GenAI use in higher education (Batista et al., 2024), and Ariza et al. examined its use in general engineering education (Ariza et al., 2025). Others are more targeted to computational thinking, but more limited in scope, with Cambaz providing a short review on programming instruction (Cambaz and Zhang, 2024), and Garcia et al. conducting a rapid review on computer programming focus on programming instruction (Garcia and B, 2025).

It should be noted that in our contribution, we do not aim to evaluate the effectiveness of GenAI learning experiences based on outcomes. Given the limited and uneven quality of empirical evidence in this emerging field (Weidlich and Gašević, 2025), the causal implications of some of the reviews that have attempted to do so are contested, as their methodologies make it hard to assess the quality of the primary studies they map out (Weidlich and Gašević, 2025). In contrast, the present review aims to map the current landscape of GenAI use in CT education. It focuses on the types of solutions developed, the contexts in which they were deployed, their evaluation maturity, how learners engaged with them, and the learning outcomes reported. It

also offers structured insights into their limitations and provides guidelines for effective use.

## 2. Methodology

The study follows a structured systematic scoping review process, based on PRISMA guidelines (Takkouche and Norman, 2011) to identify and analyze research at the intersection of GenAI and CT, selecting studies that report on learning outcomes. The methodology consists of four main phases: literature search, screening filter, eligibility evaluation, and analysis.

### 2.1. Literature Search

To ensure comprehensive coverage of relevant studies, we conducted a systematic literature search across six academic databases: ACM Digital Library, AIS eLibrary, DBLP, IEEE Xplore, EBSCOHost, and Scopus. The search used each database's default settings. In ACM and AIS eLibrary, queries targeted the full text. In IEEE Xplore, they covered metadata fields such as abstracts, indexing terms, and bibliographic data. In EBSCOHost, the search included title, abstract, subject headings, and keywords. In Scopus, it was limited to title, abstract, and keywords. The search query was formulated using three main categories of keywords: (1) Generative AI-related terms, (2) Computational thinking-related terms, and (3) Learning-related terms. Table 1 presents the terms used for each category. The query was applied to all selected databases, filtering results to include peer-reviewed journal articles and conference papers. The search was conducted in October 2024 without any filter on the publication year and yielded a total of 1,198 articles.

Category	Search Terms
<b>Generative AI</b>	Generative AI, Large Language Model, ChatGPT, Copilot
<b>Computational Thinking</b>	Computational Thinking, Algorithm Design, Problem Decomposition, Pattern Recognition, Abstraction, Programming, Coding, Software Development
<b>Learning</b>	Learning Outcomes, Student Performance, Academic Achievement, Educational Assessment, Assessment Scores, Exam Performance, Grades

Table 1: Search terms used in the systematic literature review

### 2.2. Screening Process

After removing duplicates and records not representing individual articles (e.g., full proceedings volumes), 633 articles remained for title and abstract screening. During this stage, we excluded 576 records that did not include a combination of GenAI use, CT concepts, and learning outcomes. This ensured our selection aligned with the study’s scope, which centers on the intersection of these three elements. Finally, we sought to retrieve the full text of the remaining 57 articles. Of these, three were excluded due to unavailability ( $n = 1$ ) or because they were not peer reviewed ( $n = 2$ ).

### 2.3. Eligibility Assessment and Inclusion

A total of 56 reports underwent full-text eligibility assessment. Selections were further refined by excluding articles that were not written in English ( $n = 1$ ), consisted only of an abstract with no full content available ( $n = 4$ ), or did not report on learning outcomes despite covering GenAI and CT concepts ( $n = 11$ ). Ultimately, 38 studies met all inclusion criteria and were included in the final synthesis. These studies form the basis of our analysis

on the landscape of GenAI use in the context of CT to support learning outcomes. Figure 1 shows a PRISMA flow diagram depicting the selection process in full.

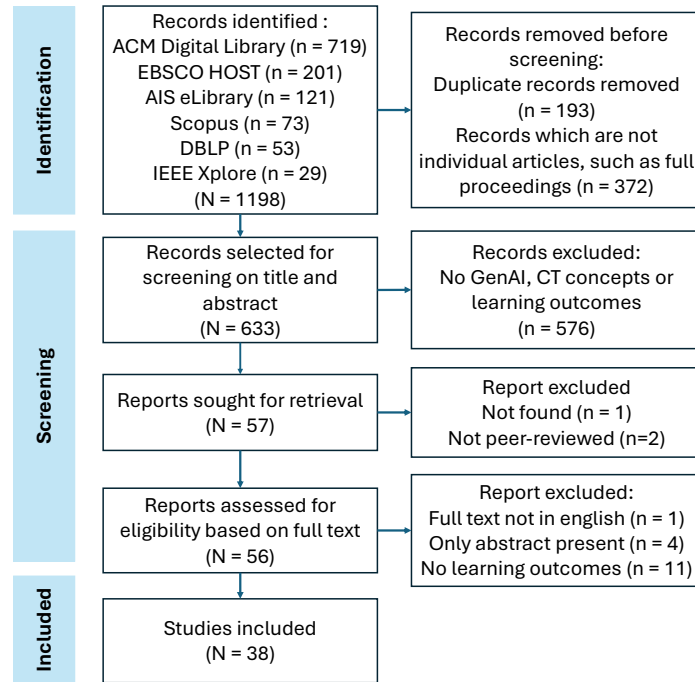


Figure 1: Selection process of papers for the literature review

#### 2.4. Analysis Procedure

The selected papers were analyzed using a concept-centric approach (Webster and Watson). A coding framework was developed to capture key aspects of how GenAI is used in teaching and learning computational thinking.

The framework included several dimensions, such as the learning context (who the learners were, course types, and learning activities), how AI was integrated (from simple tools to fully embedded solutions), and the different

roles AI played (such as coding assistant, tutor, or idea generator). The research designs and evaluation methods used in each study were also coded, covering approaches like experiments, within-subject studies, and observations.

The coding process was iterative and performed by two coders. As the analysis progressed, categories were added or adjusted to better reflect the variety of approaches and designs found in the studies. Any unclear cases or overlaps were discussed and resolved to ensure consistency. Once coding was complete, the data were used to build a synthesis that helped organize the studies, compare them, and prepare for identifying trends and research gaps in the next steps of the review. The synthesis phase integrated these coded elements to identify key research trends, recurring challenges, and avenues for future investigation, including recommendations to guide the educational use of Generative AI.

### 3. Results

In the corpus, CT concepts were primarily represented through the broader term *programming*, which appeared in 33 papers (87%). Among the core CT constructs defined by Wing and Jeannette (2011), *algorithm design* was the most frequently observed ( $n = 24$ , 63%), followed by *abstraction* ( $n = 8$ , 21%), *decomposition* ( $n = 5$ , 13%), while *pattern recognition* was only rarely mentioned ( $n = 2$ , 5%).

Algorithm design was typically operationalized through tasks requiring learners to produce executable solutions, sometimes under constraints such as passing test cases or meeting performance requirements such as memory

or time limits, e.g. Qureshi and Basit (2023), Yang et al. (2024a). Decomposition was reflected in programming activities where learners were required to break down complex problems into smaller subgoals and, when relevant, interact with GenAI systems iteratively for each part of the solution, e.g. Vadaparty et al. (2024), Liao et al. (2024). Abstraction was operationalized through tasks requiring learners to reduce problem complexity by focusing on essential features while ignoring unnecessary details in order to formulate solutions, e.g. Ouazaki et al. (2024), Liao et al. (2024). Pattern recognition, although occasionally mentioned, was not accompanied by explicit descriptions of how it was operationalized in learning tasks.

GenAI interactions were used either to engage with specific computational thinking (CT) concepts or with CT as a broader, unified construct. For algorithm design, students used GenAI agents in two main ways: either to generate an algorithmic solution to a given problem, e.g. Ouazaki et al. (2024); Qureshi and Basit (2023), or to obtain feedback and suggestions for improving algorithms they had already developed, e.g. Liao et al. (2024); Yilmaz and Yilmaz (2023). For decomposition, most studies suggest that learners remain primarily responsible for breaking down problems into sub-components before leveraging GenAI to implement these parts, e.g. Liao et al. (2024), Vadaparty et al. (2024). However, Kazemitabaar et al. (2024) note that large language models can also support the decomposition process itself. For abstraction and pattern recognition, the corpus provides limited explicit evidence on how GenAI supports these specific CT dimensions, as they are generally addressed implicitly within broader computational thinking activities rather than as distinct, targeted interactions.

In the following, we present a detailed analysis of the corpus structured along our four research questions, focusing on context, solution, evaluation, along with risks and recommendations. Figure 2 provides an overview of the distribution of the corpus per year and publisher. Given the recency of the topic, the studies were mainly published between 2023 and 2024. Table 2 shows an overview of all results.

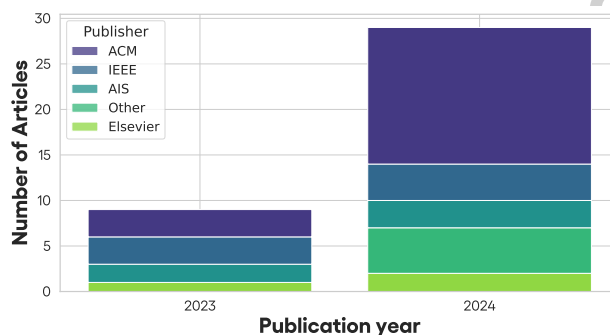


Figure 2: Distribution of publications in the corpus per year and publisher

### 3.1. Learning Context

To answer RQ1, we provide descriptions of the learning contexts discussed in the reviewed literature.

#### 3.1.1. Learners

The majority of studies ( $n = 28$ , 74%) focused on undergraduate students, followed by master's students ( $n = 5$ , 13%). Only three studies (8%) involved younger learners, including one study that focused on high school students ((Abolnejadian et al., 2024)), and another that recruited even younger participants. The latter recruited 69 youth aged 10-17 ( $M = 12.4$ ,  $SD = 1.8$ ) from

Table 2: Categorization of studies.

**Legend:** Context — Learning Experience (LX): 🧑 = Individual, 👥 = Group; Course Level: ○ = General, ● = Introductory, ● = Advanced; Age: 🧒 = Teen, 🧑 = Undergraduate, 🎓 = Graduate. Solution — Interaction (UI): 🖱️ = Click, 🗑️ = Conversational; Integration: ●○○ = Basic, ●●○ = Intermediate, ●●● = Advanced. AI Role: <math>\langle \rangle</math> = Coder, 🧑 = Tutor, 🧑 = Debugger, 🧑 = Ideator, (-) = Not reported. Evaluation — Study Design: ✂️ = RCT, ⇄ = Quasi-Experimental, ○ = Within-Subject, ● = Observational; Outcomes: green (↑) = positive, red (↓) = negative, yellow (=) = no change/mixed, gray (-) = measured but not reported.

Paper	Context				Solution				Evaluation			
	LX	Lev	Age	UI	Integ.	AI Role	Study	LO	UX	Beh.	Att.	Mot.
(Katavic et al., 2023)	-	○	🧑	-	-	-	-	↑	-	↑	-	-
(Qi et al., 2024)	-	○	🧑	-	-	-	●	↑	-	↑	-	-
(Groothuisen et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	○	↑	-	↑	-	-
(Samarakoon et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	-	○	↑	↑	↑	-	-
(Santos and Cury, 2023)	🧑, 🧑	○	🧑	🖱️	●○○	-	⇄	↑	↑	↑	-	-
(Qureshi and Basit, 2023)	🧑, 🧑	○	🧑	🖱️	●○○	-	○	↑	↑	↑	↑	-
(Khatib and Mattalo, 2024)	🧑, 🧑, 🧑	○	🧑	🖱️	●○○	-	✂️	↑	↑	↑	↑	-
(Li et al., 2024)	🧑, 🧑	○	🧑	-	●○○	<math>\langle \rangle</math>	⇄	↑	↑	↑	↑	-
(Xue et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>	⇄	↑	↑	↑	↑	-
(Yilmaz and Yilmaz, 2023)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	⇄	↑	↑	↑	↑	-
(Jošt et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	○	↓	↑	↑	↑	-
(Ouazaki et al., 2023)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	○	↑	↑	↑	↑	-
(Margulieux et al., 2024)	🧑, -	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	-	↑	↑	↑	↑	-
(Huesca et al., 2024)	🧑, 🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	✂️	↑	↑	↑	↑	-
(Hu et al., 2023)	🧑, 🧑	○	🧑	🖱️	●○○	-	○	↑	↑	↑	↑	-
(Vadaparty et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>	○	↑	↑	↑	↑	-
(Kazemitabaar et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	○	↑	↑	↑	↑	-
(Chang and Chien, 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>	○	↑	↑	↑	↑	-
(Yang et al., 2024a)	🧑, 🧑	○	🧑	🖱️	●○○	-	✂️	↑	↑	↑	↑	-
(Frankford et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	🧑	○	↑	↑	↑	↑	-
(Gottipati et al., 2023)	🧑, 🧑	○	🧑	🖱️	●○○	🧑, 🧑	○	↑	↑	↑	↑	-
(Ohm et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	-	⇄	↑	↑	↑	↑	-
(Gang and Rajendran, 2024)	🧑, 🧑	○	🧑	🖱️	●○○	-	⇄	↑	↑	↑	↑	-
(Padiyath et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑	○	↑	↑	↑	↑	-
(Kosar et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑	✂️	↑	↑	↑	↑	-
(Peixoto et al., 2024)	🧑, 🧑, 🧑, 🧑	○	🧑	🖱️	●○○	-	○	↑	↑	↑	↑	-
(Hsin and Wen-Jung, 2024)	-	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	○	↑	↑	↑	↑	-
(Rajala et al., 2023)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑, 🧑	○	↑	↑	↑	↑	-
(Mezzaro et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	○	↑	↑	↑	↑	-
(Ouaazki et al., 2024)	🧑, 🧑, 🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	○	↑	↑	↑	↑	-
(Sheese et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	⇄	↑	↑	↑	↑	-
(Liao et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	⇄	↑	↑	↑	↑	-
(Sengewald et al., 2024)	🧑, 🧑	○	🧑	🖱️	●○○	🧑	⇄	↑	↑	↑	↑	-
(Yang et al., 2024b)	🧑, 🧑	○	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	⇄	↑	↑	↑	↑	-
(Lyu et al., 2024)	🧑, 🧑	-	🧑	🖱️	●○○	<math>\langle \rangle</math>, 🧑, 🧑	⇄	↑	↑	↑	↑	-
(Kazemitabaar et al., 2023)	🧑, 🧑	-	🧑	🖱️	●○○	-	⇄	↑	↑	↑	↑	-
(Abolnejadian et al., 2024)	🧑, 🧑	-	🧑	🖱️	●○○	-	⇄	↑	↑	↑	↑	-
(Gabriella et al., 2024)	-	-	🧑	🖱️	●○○	-	⇄	↑	↑	↑	↑	-

participants of coding camps in two cities, and resulted in two publications (Kazemitabaar et al., 2024, 2023).

### 3.1.2. Course subject

The studies reviewed covered a variety of courses dedicated to improving computational thinking skills. These courses can be grouped into three key thematic categories: general introduction, basic programming, and advanced programming. General introduction studies (8 papers, 21%) were conducted in general IT or mathematics courses, where computational thinking skills were only peripheral to the main content. More specifically, 3 studies (8%) focused on *mathematics*, algebra, and scientific computing, while 5 studies (13%) focused on *general introductory courses*, such as Information Technology, Digital Business, or AI in Everyday Life. Furthermore, multiple studies ( $n = 12$ , 32%) were conducted in courses that explicitly teach basic programming. Finally, a large proportion of papers ( $n = 14$ , 37%) covered more advanced or specialized topics, such as object-oriented programming, web development, networking, cybersecurity, and software design. Several of these advanced courses were explicitly about *data science*, covering Data Analysis in Engineering, Data-Oriented Programming, Data Structures and Algorithms, as well as database courses such as SQL.

Figure 3 visualizes the learning contexts covered with a heatmap showing the educational levels of learners and types of courses. It shows that most studies target undergraduate learners, especially in advanced courses, with basic and general courses also well represented, while teens and graduate students are less frequently addressed. Note that three studies provide learning experiences that are not targeted at a specific course and as such are not

represented in Figure 3.

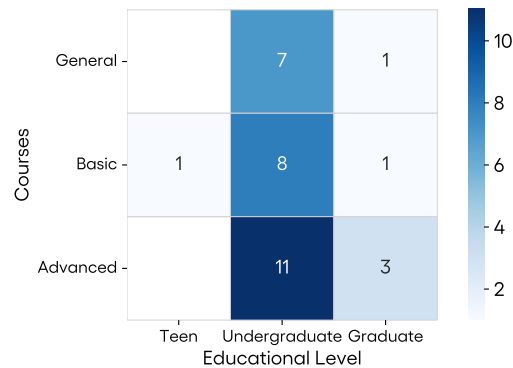


Figure 3: Studies mapped by Educational Level and Course Type

### 3.1.3. Learning Experiences

The reviewed studies implemented a variety of learning experiences. As shown in Figure 4, most studies focused on individual learning, where learners completed assignments independently ( $n = 26, 68\%$ ). Eleven studies (29%) targeted *collaborative learning*, involving two or more students working together. Some of these were explicitly structured as *pair programming*, where two learners work together at one computer, one writing code while the other reviews and guides, and *peer instruction*, where learners explain concepts to one another to reinforce understanding. While most papers explored the use of GenAI in traditional learning classrooms, 2 studies (5%) described interventions in flipped classrooms, where students prepare content before class and consolidate their knowledge through in-class activities.

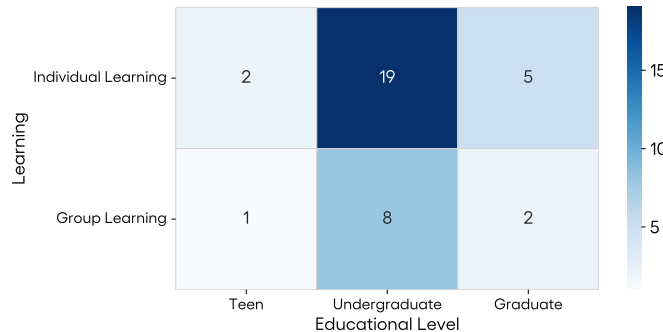


Figure 4: Studies mapped by Educational Level and Learning Experience

### 3.2. GenAI Solutions

To answer RQ2, we analyzed the GenAI solutions presented in the literature to support the above-mentioned learning experiences.

#### 3.2.1. Models

Figure 5 shows the main GenAI models used in the reviewed studies. The majority of studies used an OpenAI GPT model accessed directly through the ChatGPT interface ( $n = 20$ , 53%). Many studies also use OpenAI's GPT models via API access ( $n = 15$ , 40%). GitHub Copilot is used in 4 studies (11%), while other studies reported using OpenAI's Codex ( $n = 2$ , 5%), or EleutherAI's GPT-J-6B ( $n = 1$ , 3%). It is worth noting that model counts are not mutually exclusive, as a single paper in our coding could include more than one GenAI model. The concentration of OpenAI-based products in the reviewed studies is not surprising, as OpenAI's GPT models, whether accessed through the ChatGPT interface or via API, were among the first widely adopted GenAI tools in education (Baig and Yadegaridehkordi, 2024). As a result, the current evidence base is largely shaped by OpenAI-

based systems rather than reflecting the full diversity of emerging GenAI tools, including Claude and Gemini, that are increasingly gaining traction. We also acknowledge recent developments in agentic AI systems in education Kostopoulos et al. (2025), in which models can autonomously plan, execute, and iterate on multi-step tasks; however, these fall outside the scope of the corpus considered in this review, given the search timeline.

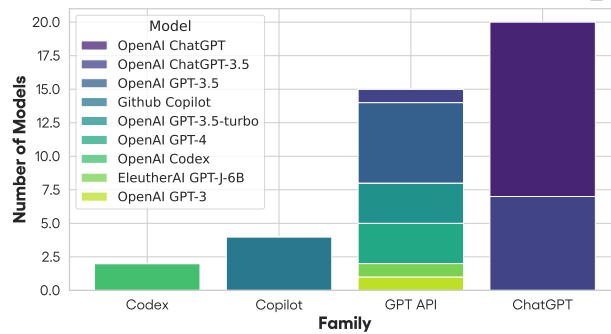


Figure 5: Distribution of LLM types used in the studies

### 3.2.2. Learning experience integration

In terms of integration with the learning experience, there have been three levels of integration within the curriculum reported in the studies.

*Level 1 - Tool-as-is (No contextual integration).* On the first level, many studies simply involved allowing students to use off-the-shelf GenAI tools (mainly ChatGPT) as a support ( $n = 21$ , 55%). Sometimes, students were given a prompt that they can use in the system ( $n = 5$ , 13%). At this level, the students had to choose the GenAI tool themselves and provide it with all necessary information through user prompts, as the tool had no awareness of the learning context.

*Level 2 - Basic platform integration.* At the second level, some studies integrated a GenAI tool into a learning system used in the course ( $n = 15$ , 39%), typically by adding a GenAI-enabled conversational agent to an existing interface. These agents were prompt-tuned in a basic way, receiving general behavioral instructions and limited contextual information. For example, in (Rajala et al., 2023), a chatbot named Kiran was embedded in Microsoft Teams, allowing students to ask curriculum-related questions. At the first level, and in all but two studies at the second level, user interaction with GenAI took place through open, unrestricted chat conversations. The remaining two studies at the second level restricted interactions to predefined options, such as button clicks, rather than allowing free-text input.

*Level 3 - Advanced platform integration.* At the third level, the GenAI tool is fully integrated into the system, with access to the assignment context, students' live code, and the reference solutions ( $n = 9$ , 24%). As such, the system could act as a tutor, as illustrated in the prompt shown in Figure 6. At this level, the model could also be personalized according to the specific background of the student, including their age, location, grade, and IQ (Abolnejadian et al., 2024) as shown in Figure 7. In this particular study targeting teens, and in several other fully integrated solutions ( $n = 6$  in total, 16%), interaction with the GenAI tool was limited to simple clicks rather than open-ended input.

Figure 8 presents a heatmap showing the distribution of studies across different levels of GenAI integration and educational levels of learners. It shows that most studies focus on undergraduate students with Level 1 integration ( $n = 16$ , 42%). Additionally, Figure 9 shows that project-based

```
Act as a programming tutor and give informal feedback
in [language] to the student.
```

```
The exercise description is the following:
[description]
```

```
The students code looks like that at the moment:
[current code]
```

```
Do not provide a code solution.
```

```
The optimal solution should look like that:
[solution]
```

```
Important: Do not provide code.
```

Figure 6: Example of a full integration fine-tuned prompt by (Frankford et al., 2024).

```
Act as an introductory programming teacher who teaches Python.
I am a AGE year old student in GRADE grade from a school with SCHOOL
utilities and a CITY prestige city.
My GPA is GPA out of 20, and my math score is MATH out of 20.
I got an IQ out of 100 scores on an IQ test.
I PROGRAMMING.
Tailor your answers to the specific background that I provided.
```

Figure 7: Example of a full integration fine-tuned prompt with personalization by Abolnejadian et al. (2024).

studies did not implement any integration of Level 3, and only two studies have Level 2 integration. One of these two studies investigated Copilot within VScode (Vadaparty et al., 2024), while the other used the integration of Level 2 for its exercise-based assignment and integration of Level 1 for its

project-based assignment (Ouaazki et al., 2024).

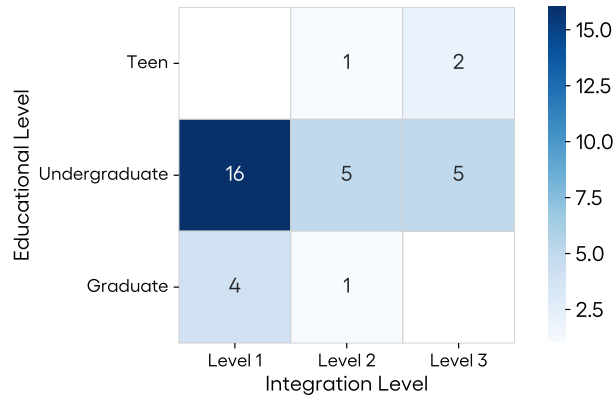


Figure 8: Studies mapped by Integration and Educational Levels

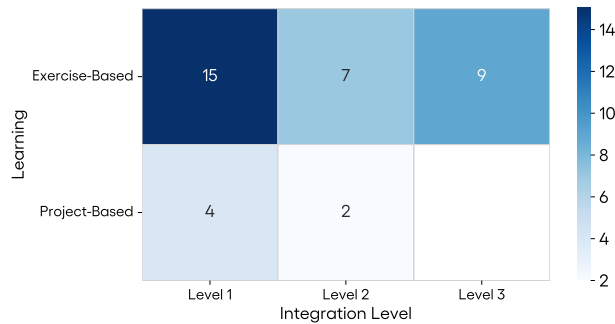


Figure 9: Studies mapped by Integration level and Learning Experience

### 3.2.3. GenAI Role

Across the reviewed literature, many papers ( $n = 25$ , 66%) explicitly described how GenAI tools supported learners in their computational thinking processes. We inductively categorized these uses into four main roles: *Coder*, *Tutor*, *Debugger*, and *Ideator*. These categories capture the diverse ways learners engaged with GenAI, ranging from conceptual support and

error correction to code generation and idea exploration. In the following paragraphs, we describe each category in detail.

*Coder.* Using GenAI as a *coder* that performs implementation tasks was the most frequently reported category, mentioned across 22 papers (58%). Learners engaged with GenAI primarily to transform their ideas into executable programs (e.g., (Yang et al., 2024b; Xue et al., 2024; Vadaparty et al., 2024; Ouazaki et al., 2024)). In many cases, this involved generating new code or assembling complete solutions with the assistance of the tool. GenAI also supported learners in modifying existing code, enabling them to adapt or refine outputs based on evolving needs (e.g., (Lyu et al., 2024; Kazemitabaar et al., 2024)). Some papers reported the use of GenAI for optimizing code, helping learners improve efficiency or structure (e.g., (Groothuisen et al., 2024; Kosar et al., 2024)). GenAI assisted in solving individual components or multiple parts of a larger task, guiding learners through the step-by-step development of a complete solution (e.g., (Margulieux et al., 2024)).

*Tutor.* GenAI was used as a *tutor* that helps learners understand the material at hand in 19 papers (50%). Learners used GenAI to deepen their grasp of programming concepts, clarify problem statements, and make sense of unfamiliar code. Interactions often involved asking for explanations, seeking feedback, or comparing different approaches to enhance comprehension. For instance, (Gottipati et al., 2023) illustrated an *understanding* use through a sensemaking interaction, where learners engaged in decision-making informed by ChatGPT's responses. Similarly, (Rajala et al., 2023) reported that the tool provided new perspectives and tutoring support, helping learners deepen their subject understanding. GenAI was also used to support conceptual

understanding by helping learners test hypotheses and explore alternative solutions (e.g., (Yang et al., 2024b)). GenAI tools also supported learners in understanding how to effectively decompose problems, guiding their efforts toward structured and manageable solutions (e.g., (Liao et al., 2024)).

*Debugger.* Debugging tasks were another common use case of GenAI, with 11 coded instances across 11 papers (29%). Learners turned to GenAI to identify and correct errors in their code, benefiting from the tool's ability to analyze outputs, suggest fixes, and explain potential issues (e.g., (Kazemitabaar et al., 2024; Jošt et al., 2024; Padiyath et al., 2024)). GenAI supported both the detection of bugs and the repair of faulty logic or syntax (e.g., (Ouaazki et al., 2024; Kazemitabaar et al., 2024)).

*Ideator.* GenAI was used as an *ideator* that helped learners with creative tasks for CT across 5 papers (13%). It reflects how GenAI supported learners in generating original ideas, exploring alternative solutions, and expanding their problem-solving approaches. Several studies highlighted GenAI's role in helping learners formulate new directions or subgoals during task completion, encouraging a more open-ended and exploratory mindset (e.g., (Rajala et al., 2023; Kazemitabaar et al., 2024)). Other examples emphasized learners' ability to ask novel questions and pursue unconventional strategies with the support of the tool (e.g., (Yilmaz and Yilmaz, 2023)). In some cases, creativity emerged through reciprocal prompting, where the learner and GenAI co-constructed new ideas or directions (e.g., (Katavic et al., 2023)). GenAI was also used to support creativity in how learners approached and structured their problem-solving processes, such as *wayfinding*, where learners interacted with ChatGPT to create a knowledge pipeline by asking questions,

and innovation in exploring new directions (e.g., (Gottipati et al., 2023)).

### 3.3. Evaluation

To answer RQ3, we analyzed the different empirical evaluation methods and results found in the literature. In the following, we discuss the study types and metrics used.

#### 3.3.1. Study Types

The reviewed studies employed various research methodologies, including randomized controlled trials (RCTs), quasi-experimental designs, and observational studies. It should be noted that all but two evaluation setups have taken place in authentic educational settings.

*Randomized controlled trials.* Multiple studies ( $n = 6$ , 16%) implemented RCTs to test the impact of GenAI based on a control and experimental condition where participants are randomly assigned. This design allowed investigating the effect of the interventions in controlled conditions, minimizing the effects of external variables. For instance, in (Yilmaz and Yilmaz, 2023), participants were randomly divided into a control and experimental group completing activities such as weekly hands-on Java programming assignments (with or without ChatGPT support), to test their differences in computational thinking, computer programming self-efficacy, and learning motivation.

*Quasi-experimental designs.* Many studies ( $n = 8$ , 21%) implemented interventions using quasi-experimental designs, with a control and an experimental group that were not randomly assigned, but assigned based on a set of

predefined rules. For instance, in (Kazemitabaar et al., 2023), participants were divided into two groups using a matched-groups design, ensuring that both groups had similar means and variances based on their pre-test scores on a programming assignment.

*Within-subject experiments.* A large body of studies ( $n = 17$ , 45%) was based on field experiments using a within-subject design studying the impact of GenAI use on participants. For example, in (Yang et al., 2024b), students were interviewed and participated in think-aloud debugging tasks at three points throughout the semester, allowing researchers to track changes in help-seeking behavior and perceived learning over time.

*Observational studies without intervention.* This category includes studies that collect data without implementing an intervention ( $n = 5$ , 13%). It is based on surveys distributed to participants to collect quantitative data, or interviews to gather qualitative insights. These studies aim to describe the effects of GenAI on CT skills without manipulating variables. For instance, (Padiyath et al., 2024) surveyed students after coursework to examine their attitudes and use of LLMs, followed by interviews to explore adoption contexts and experiences with AI tools like ChatGPT in programming classes.

### 3.3.2. *Collected Measures and Outcomes*

Many measures have been used across different studies. Most data were collected through custom assignments and surveys, along with additional methods such as logs, eye-tracking, and screen recordings. We have categorized them into four categories, which we detail below. We also examined the specific outcomes associated with each measure, coding whether the pa-

pers reported a positive impact, a negative impact, no difference, or did not measure/mention an effect. The results of this coding are shown in Table 2.

*Learning outcomes.* Metrics related to learning outcomes relevant to CT were collected in all reviewed articles. The learning outcomes mostly assessed student performance through tests, assignments and exams ( $n = 30$ , 79%). These tests targeted measures such as accuracy scores, completion rates and time spent on exercises. For example, Ohm et al. (2024) measured learning outcomes through weekly assignments and final exam grades, along with transfer scores that assessed the learner's capacity to apply acquired knowledge to novel problems or contexts.

Additionally, complementary aspects of learning such as perceived computational thinking skills, engagement, confidence and creativity were assessed using self-reported measurements ( $n = 12$ , 32%). These were evaluated either using custom questionnaires made for the specific context of the study ( $n = 9$ , 24%) or standard scales ( $n = 4$ , 11%) such as the Computational Thinking scale (Tsai et al., 2021), and the Interaction and Cognitive Engagement in Connectivist Learning scale Wang et al. (2014). For instance, in (Yilmaz and Yilmaz, 2023) and (Liao et al., 2024), learning was assessed using the Computational Thinking Skills scale, which focuses on algorithmic thinking, creativity, and critical thinking. This diversity of measurement approaches highlights the lack of standardized methods for assessing computational thinking in GenAI-supported learning contexts, calling for greater standardization in future research.

Overall, 25 papers (66%) reported improvements in learning outcomes including performance measures such as exam scores, completion and submis-

sion rates, and self-reported measures such as perceived learning, confidence, creativity and computational thinking levels. Six papers (16%) reported decreases in learning outcomes such as final exam grades and transfer and reproduction skills. Finally, 6 papers (16%) reported no change in measures like weekly assignment grades, submission rates or creativity. It is worth noting that some papers reported increases in certain outcomes while observing a decrease in others. These mixed results should be interpreted in light of the methodological diversity of the corpus. A total of 13 studies (35%) reported some form of randomization, with only 6 studies (16%) implementing RCTs. Sample sizes varied across studies, ranging from 12 (Hsin and Wen-Jung, 2024) to 400 (Gabriella et al., 2024) (median  $\approx 50$ ). Power analyses were not reported in any of these studies.

*User experience (UX).* Several studies ( $n = 15$ , 39%) examined UX dimensions to understand how students perceived and interacted with GenAI tools. Reported metrics included perceived usefulness, ease of use, satisfaction, usability, and trust. UX was primarily assessed through self-reported questionnaires and interviews. Some studies applied established models, such as the Technology Acceptance Model (Davis et al., 1989) to evaluate perceived usefulness and ease of use (Frankford et al., 2024; Khatib and Mattalo, 2024), and the Learning Management System Acceptance Scale (Sezer and Yilmaz, 2019) to assess overall usefulness and user perceptions (Li et al., 2024). Others relied on custom scales to measure satisfaction, GenAI assistance, and usability (Abolnejadian et al., 2024). Additional studies also addressed concerns such as fear of surveillance and trust in GenAI (Rajala et al., 2023). Among the reviewed papers, 8 studies (21%) reported improvements in UX,

2 (5%) reported decreases, and 2 (5%) reported no change.

*Behavioral outcomes.* Several studies ( $n = 11$ , 29%) also examined behavioral outcomes to capture how GenAI tools influenced students' actions and strategies during learning. Reported metrics included collaboration between students, attention level, intention to use, engagement, and self-efficacy, which were often measured through self-reported surveys. Self-efficacy, for example, was a common focus and was investigated in multiple studies (Qi et al., 2024; Yilmaz and Yilmaz, 2023; Padiyath et al., 2024; Margulieux et al., 2024; Gabriella et al., 2024), each exploring how confident students felt while using GenAI. Collaboration was another key outcome, examined through measures of cooperation in studies by Yilmaz and Yilmaz (2023) and Liao et al. (2024). Overall, 5 studies reported positive effects on behavioral outcomes such as engagement, self-efficacy, and intention to use, 2 studies reported negative effects, particularly on self-efficacy and collaboration, and 2 reported no significant change in measures like cooperativity and self-regulation.

*Attitudinal outcomes.* Several studies ( $n = 7$ , 18%) investigated attitudinal outcomes to understand students' perceptions and dispositions toward GenAI tools. Reported metrics included general attitudes toward technology, specific attitudes toward ChatGPT and other AI systems, and comfort with using generative AI. For example, attitudes toward ChatGPT were explored in Xue et al. (2024), while comfort levels with GenAI were assessed using custom questionnaires in studies such as Vadaparty et al. (2024). In total, 4 studies (11%) reported more positive attitudes, none reported more negative attitudes, and 3 (8%) reported no change following GenAI use.

*Motivational outcomes.* Motivational outcomes were less frequently measured in the reviewed studies ( $n = 2$ , 5%). For example, Yilmaz and Yilmaz (2023) examined motivation through factors such as reward and recognition, punishment, and social pressure or competition, using the Learning Motivation in Computer Programming Courses Scale (Law et al., 2010). Similarly, Gabriella et al. (2024) focused on learning motivation to assess how generative AI influenced students' drive to engage with course material. In total, only one study reported an increased motivation.

### 3.4. Limitations and Recommendations

To answer RQ4, we present the main limitations associated with the use of GenAI learning experiences reported in the literature, along with recommendations for optimal GenAI use.

#### 3.4.1. Limitations

In terms of limitations, we have identified two categories: GenAI-related limitations, which are related to the tool itself, such as inaccuracies and reliability issues, and learner-related limitations, which limit learners, such as over-reliance and threats to critical thinking. In total, 20 papers (53%) mentioned some limitations of GenAI.

*GenAI inaccuracies and reliability issues.* Multiple studies ( $n = 9$ , 24%) reported concerns related to the accuracy and reliability of GenAI tools. Some studies highlighted uncertainty and inconsistency in ChatGPT's outputs, as well as limited transparency regarding its underlying mechanisms (Rajala et al., 2023). Others observed that the tool often failed to produce accurate solutions or generated code that did not pass test cases, raising doubts about

its proficiency (Qureshi and Basit, 2023). In some cases, the production of non-optimized or irrelevant code led students to abandon ChatGPT and return to traditional learning resources (Hu et al., 2023), or sometimes failed to fully grasp questions or contextual details (Qureshi and Basit, 2023; Rajala et al., 2023).

Studies reported that these shortcomings sometimes impacted students' perceptions of the tools' usefulness, which tended to decline over time (Margulieux et al., 2024). This, in turn, led to reduced reliance on the tool, especially as task complexity increased (Ouaazki et al., 2024), and resulted in some students turning back to human teaching assistants for support (Lyu et al., 2024).

*Overreliance and threats to critical thinking.* While some students were critical of GenAI and even abandoned its use due to its shortcomings, many researchers (n = 17, 44%) observed that students often failed to recognize these limitations. As a result, they became overly reliant on GenAI, which not only produced inadequate solutions but also hindered learning, leading to dependency (Liao et al., 2024; Ouaazki et al., 2024) and reduced performance when GenAI was no longer available (Ohm et al., 2024). Indeed, an overarching concern around the use of GenAI in education is its potential to trigger “metacognitive laziness” that can hinder students' ability to foster critical thinking (Fan et al., 2025; Lyu et al., 2024) and engage deeply with their learning through self-regulation (Xue et al., 2024; Jošt et al., 2024).

### 3.4.2. Recommendations

Overall, 27 papers (71%) presented recommendations about GenAI use, emphasizing the need to train learners in prompting, foster GenAI literacy, and adapt CT learning experiences to account for GenAI.

*Train learners in prompting.* Many studies ( $n = 11$ , 29%) highlighted the importance of teaching students how to write effective prompts when working with GenAI tools to support their development of programming and CT skills. Students themselves reported that asking the right questions was key to obtaining accurate and detailed answers from ChatGPT, yet many lacked the skills to frame questions properly (Liao et al., 2024). Prompt writing was highlighted as essential for effective AI use (Yilmaz and Yilmaz, 2023), with studies also noting that precise prompts are crucial for generating useful outputs, similar to search engines (Qureshi and Basit, 2023). Several approaches were proposed to improve students' prompting skills, including prompt construction templates (Lyu et al., 2024), interactive dialogue-based systems promoting task decomposition (Kazemitabaar et al., 2023), and iterative prompting, which proved more effective than single-prompt approaches (Kazemitabaar et al., 2024).

*Foster GenAI literacy.* Multiple studies ( $n = 9$ , 24%) also emphasized the importance of educating students and providing them with proper guidance when integrating GenAI into CT education. Students themselves expressed the need for introductory tutoring to better understand ChatGPT and the principles behind large language models (Rajala et al., 2023). To support this, some studies recommended that instructors offer clear guidelines on the

appropriate use of ChatGPT, such as facilitating discussions to familiarize students with its capabilities and limitations (Xue et al., 2024), rather than letting students figure them out in the wild. Further, it was argued that providing students with training or guidance could improve learning outcomes by helping them set realistic expectations, avoid misuse, and address tool limitations (Mezzaro et al., 2024). At the curricular level, the integration of GenAI literacy as a core component was suggested to help students effectively use these tools for problem-solving (Lyu et al., 2024).

*Adapt learning experiences.* Almost half of the reviewed studies ( $n = 18$ , 47%) emphasized the need to adjust learning activities to account for the use of GenAI tools. At a basic level, some studies suggested refining responses to deliver more detailed, code-specific guidance to better support learners (Frankford et al., 2024) - potentially tailored to students' individual needs to ensure feedback remains valuable as they progress (Yang et al., 2024a; Sheese et al., 2024). Some researchers noted that traditional assignments such as quizzes may no longer be effective, as students can easily solve them by simply asking ChatGPT or similar GenAI tools (Yilmaz and Yilmaz, 2023). To address this, research suggested that assessments should focus on students' ability to apply CT concepts, rather than on questions that can be answered through straightforward information retrieval (Qureshi and Basit, 2023). In this context, designing complex programming tasks remains valuable, as they still require students to break down problems and generate code independently (Groothuijsen et al., 2024). Further recommendations include integrating assignment defenses, i.e., oral presentations where students defend their assignments in front of an instructor through interviews

and discussions, to encourage deeper engagement with the material (Kosar et al., 2024). Some researchers argued that to make effective use of AI tools, it is essential to foster students' thinking skills and imagination (Yilmaz and Yilmaz, 2023). Maintaining a balance between support and autonomy was also viewed as critical. For instance, providing assistance while ensuring that learners retain responsibility for their own progress was seen as key to effective AI-supported learning (Yang et al., 2024a), which also applies to CT skills. To promote higher-level thinking, several techniques were proposed. One approach involves using GenAI to generate sub-programs and then critically integrating them into a larger solution (Hu et al., 2023). Another technique is embedding "Critical Thinking Prompts" into AI interfaces (Lyu et al., 2024). These prompts are designed to encourage students to think critically about the AI's output by evaluating whether responses are meaningful, accurate, and syntactically correct. A third approach applies "metacognitive prompts" to support self-reflection and learning regulation (Yilmaz and Yilmaz, 2023). In this case, students are encouraged to reflect on how they used ChatGPT during programming and what they learned from the experience.

*Establish usage policies.* Several papers ( $n = 6$ , 16%) argued for establishing GenAI policies, to cover usage (Rajala et al., 2023), misconduct (Qureshi and Basit, 2023), and shared best practices among institutions (Xue et al., 2024). Some researchers have argued that students should be free to use GenAI, as evidence points to improvements in learning outcomes, including problem-solving skills (Sengewald et al., 2024), sometimes comparing GenAI to calculators, which have become standard tools in classrooms, despite initially meeting resistance (Ohm et al., 2024). However, not all studies support

unrestricted use, as one study warned that ChatGPT may not be suitable for learning programming and recommended hands-on practice based on textbooks and online courses (Qureshi and Basit, 2023). Some have suggested that the use of GenAI tools should be postponed until students have first developed basic programming skills (Vadaparty et al., 2024; Jošt et al., 2024). Additionally, the need for clear policies to regulate the use of GenAI in education was highlighted in several studies ( $n = 4$ , 11%), with some focusing on the societal level - calling for regulations and legislation governing conversational agent usage (Rajala et al., 2023) - and others on the institutional level - emphasizing the need to update academic integrity policies and honor codes to address AI tools (Qureshi and Basit, 2023). Collaboration among institutions was also deemed essential, with the sharing of experiences, best practices, and challenges considered vital for developing effective GenAI strategies and curricula (Xue et al., 2024).

#### **4. Main Findings and Guidelines**

This paper presented a systematic scoping review of the literature on GenAI-augmented learning experiences in computational thinking (CT). After an initial search result yielding 1198 results, our screening process, focusing specifically on pieces about the use of GenAI in the context of teaching CT concepts with investigation of learning outcomes, narrowed our corpus to 38 papers, which we analyzed in detail. Our results show that the literature is very young (only two years old) but already rich and growing rapidly, with several related reviews also published on related topics. Note that this review is not without limitations. It is a scoping review on an emerging

topic, and as such, more literature becomes available. Nevertheless, this review provides a useful lens on current practices of GenAI-enhanced learning experiences in CT. Our findings show that GenAI reshapes how CT skills are enacted by shifting learners' role from producing solutions to evaluating, refining, and integrating AI-generated outputs, while moving their focus toward higher-level tasks such as problem formulation and decomposition, suggesting that educational practices should adapt to support the development of these evolving skills. Below, we discuss the findings along three main design topics that emerged from our analysis: A) Designing for simple coding exercises, B) Designing for open-ended projects, and C) Designing for awareness and reflection. For each topic we provide actionable guidelines for researchers and practitioners informed by CT as an educational framework. These guidelines were iteratively refined by the authors and contextualized in relation to the broader body of related work.

#### *4.1. Designing for simple coding exercises*

The results show that approximately two-thirds of the studies were conducted in general introductory courses or basic programming courses. In terms of learning experiences, a large majority of papers ( $> 80\%$ ) focused on simple exercises as opposed to open ended projects. Regarding GenAI integration into the learning experience, most research examined as-is integration, where students were simply given access to GenAI tools such as ChatGPT or Copilot with little or no contextual framing. In such settings, GenAI can easily solve simple exercises, such as writing a function to traverse a list and count the number of occurrences of an element, by simply copy-pasting the instructions. This ease of use and access can be tempting for

students to over-rely on GenAI, leading to their failure to critically engage with GenAI tools for deeper learning (Shibani et al., 2024). Indeed, nearly half of the reviewed papers reported that over-reliance and cognitive offloading were important issues faced by students. Some research even suggested avoiding GenAI in learning experiences for beginners at all. In the literature, two main strategies have been proposed to address this issue: (1) Limiting the capabilities of GenAI, and (2) Expanding the abilities of students.

**Guideline 1 – Guide GenAI use for beginners: Explain risks and promote tutor-style prompting:** *Support students using GenAI as-is to help them with their assignment by (1) explaining the shortcomings and risks and (2) fostering GenAI literacy by providing best practices and example prompts.*

A large majority of papers in the corpus argue for providing guidelines to students (> 70% of papers). At this stage, the literature has primarily highlighted problems like improper use, over-reliance, and cognitive laziness that emerge without guidance, while offering few concrete guidelines in response. Interestingly, even in scenarios where GenAI was used *as-is* with no contextual integration, students independently used GenAI to play different roles in their learning, from coder and debugger to tutor and ideator. For early-stage students, it can be tempting to ask GenAI to solve or debug exercises directly, but this defies the purpose of active learning and risks hindering the development of essential problem-solving skills (ElSayary and Areej, 2024). It is therefore important to raise students' awareness of these risks. Conversely, using GenAI as a tutor that answers questions without giving away the solution is an appropriate use case and can be encouraged (Ouaazki et al.,

2023; Frankford et al., 2024). Inspired by Socratic instructional strategies, this can involve a set of guiding questions that can scaffold students for self-explanations to improve their programming knowledge (Alshaikh et al., 2020). It could also be achieved through prompt construction templates, to help users craft more effective and focused queries (Lyu et al., 2024). Future research could further investigate how to best craft prompts to support students in this kind of Socratic tutoring to enhance their self-regulated learning.

**Guideline 2 – Confine GenAI interactions for novice and young learners:** *In settings involving teenagers or beginner learners, GenAI can be beneficial if used in a controlled manner, with limited interaction and response capabilities.*

Our review shows that in studies involving younger or novice learners, GenAI tools were sometimes integrated in constrained ways, such as role-limited guidance, click-based interfaces, or scaffolded prompts. In the case of role-limited guidance, this includes assigning GenAI a specific *role* with instructions on how to respond and equipping it with contextual knowledge, such as access to the current code of the student, the instructions, and the intended learning outcomes. Typically, the role of GenAI would be that of a tutor that provides guidance without revealing the answer (Ouaazki et al., 2023; Frankford et al., 2024). This allows GenAI to provide targeted support while maintaining pedagogical goals. In many beginner-level courses, such integration may take the form of click-based interactions (e.g., clicking to reveal hints or scaffolded suggestions) rather than an open-ended conversational interaction (e.g., (Chang and Chien, 2024), (Kazemitabaar et al.,

2024),(Yang et al., 2024a)). This could reduce the overload and noise that might arise from exchanging with the conversational agent (Macko et al., 2025). Designing constrained but context-aware GenAI experiences is far from trivial, but could offer greater scalability and control, particularly for younger learners. Prior studies also support limiting GenAI interactions in sensitive settings, such as with children, due to concerns of misinformation and over-trust in AI outputs (Solyst et al., 2024). Future research could further explore how to design controlled Gen-AI learning environments that include other motivational affordances, such as emotional integration and gamification, to improve learning and student engagement.

#### *4.2. Designing for open-ended projects*

The literature includes only a few studies that focus on open-ended projects. This may be because such projects typically span several weeks in authentic educational settings, making them more time-consuming and resource-intensive to study. Nevertheless, existing studies provide some valuable insights. In particular, some students find it difficult to take advantage of GenAI in such contexts where GenAI cannot solve the problem quickly. In one study, students saw the value of GenAI on their learning decrease as their project became more complex (Ouaazki et al., 2023). We define guidelines for such open-ended projects next.

**Guideline 3 – Combine GenAI with computational thinking concepts in open-ended projects:** *For open-ended projects, the roles of GenAI can be expanded beyond solution generation to include coder, debugger, and ideator functions. To be effective, students should actively engage with varying CT concepts. In current GenAI applications, the roles given to AI*

predominantly focus on algorithmic design among the computational thinking concepts, leaving aside other key components such as abstraction, finding patterns, and decomposition. Yet, these concepts are particularly helpful when dealing with more complex tasks, particularly decomposition, which allows students to solve complex problems by breaking problems into smaller sub-tasks, thereby fostering their critical thinking abilities (Rich et al., 2019). Educators should therefore design learning activities that encourage students to combine GenAI input with multiple CT strategies to enhance depth in problem-solving. Vadaparty et al. (2024) argue that educators do not need to undergo a complete course redesign to integrate GenAI effectively. Instead, essential adjustments, including teaching decomposition and testing, and using LLM-generated code in class examples, can significantly enhance learning.

**Guideline 4 – Explore advanced learning experience integration:**

*Learning experiences that incorporate GenAI in more advanced roles, such as a pair-programmer, could be leveraged to promote higher-order outcomes*

There is currently a lack of high-level (Level 3) integration for GenAI in project-based learning experiences. However, when implemented thoughtfully, these approaches could support not only technical skills but also other learning outcomes such as collaboration, creativity, and metacognitive skills (Khotimah and Rusijono, 2024). Several studies have explored GenAI as a pair programming partner, which can simulate human collaboration by suggesting alternatives, asking clarifying questions, or co-constructing solutions (Garcia and B, 2025). Future research could explore how to design learning experience through structured roles, such as assigning GenAI and learners the

roles of *driver* and *navigator* (Bryant et al., 2008), to better understand how role distribution influences learning. This line of work could offer valuable insights into effective human-AI collaboration and help design systems that support both individual accountability and collaborative problem-solving.

#### 4.3. *Designing for awareness and reflection*

The previous sections highlighted how GenAI tools can support a range of learning outcomes in computational thinking, from helping with coding tasks to fostering productivity and collaboration. Yet, these benefits are not guaranteed. As some papers indicated, the same tools could also lead to over-reliance, shallow engagement, or reduced self-regulation if not used thoughtfully. Designing effective GenAI-enhanced learning experiences therefore requires more than just tool integration, it calls for deliberate strategies that promote awareness and reflection. This includes understanding how students interact with GenAI, how these interactions shape their learning behaviors, and how educators can guide them toward more meaningful and ethical use.

**Guideline 5 – Measure outcomes for reflection:** *GenAI-enhanced learning experiences should include mechanisms that promote both learner and instructor reflection.* Our review shows that while many studies focused on learning outcomes, far fewer explored how students reflect on their use of GenAI, or how instructors adapt their teaching in response. Given the risks of overreliance and metacognitive laziness, measuring outcomes should not only serve evaluation but also inform reflective practices. This includes helping students recognize when and how GenAI supports their thinking, and encouraging them to critically assess its outputs. For example, integrating critical

thinking prompts, such as asking students whether the AI's response fully addresses their question or what additional information might be needed, can stimulate deeper analytical thinking and foster habits of reflection and evaluation (Lyu et al., 2024). For educators, systematic tracking of learning patterns and engagement can reveal when instructional adjustments are needed, for instance, by monitoring students' interactions with GenAI tools to detect inappropriate use, such as signs of plagiarism or excessive reliance, and intervening when necessary to maintain academic integrity and support deeper learning (Qureshi and Basit, 2023). Embedding reflective prompts, post-task debriefings, or learning journals into GenAI-supported activities can also strengthen self-regulation and deepen computational thinking skills. Future work could design novel scales and methods to help educators better identify both risks and the conditions for success.

**Guideline 6 – Design Learning Experiences with digital ethics and integrity in mind:** *GenAI tools can raise ethical concerns, ranging from privacy and bias, to threats to academic integrity. Identifying these issues and developing mitigation strategies is increasingly important.*

One of the most frequently reported issues is inaccuracy, so it would be useful to understand which models perform better than others. However, in the literature, the overwhelming majority of existing learning experiences use the tool of only one company, namely a flavor of OpenAI's GPT models - limiting comparative insight. Furthermore, when using OpenAI or other proprietary solutions, interaction data may be used for further model training, which can potentially pose privacy and informed consent risks, especially if students are unaware of these practices (Golda et al., 2024). A growing

body of work also highlights bias embedded in LLM outputs. A recent study has experimentally demonstrated implicit biases in large language models, showing how GenAI responses can reinforce gender and racial stereotypes, even in educational contexts (Warr et al., 2024). This underscores the importance of preparing students and educators to recognize and mitigate biased outputs, especially when these tools are used in evaluative or instructional roles. Practitioners should aim to identify these issues early and adopt mitigation strategies - from securing informed consent and model transparency, to using alternate tools and models, or modifying the learning experience. Additionally, plagiarism and academic misconduct were also recurring concerns, particularly when students used GenAI to complete assignments without understanding the underlying concepts (Hutson and James, 2024). In this context, educators have a role to play to establish clear usage policies and monitoring practices to uphold academic integrity (Qureshi and Basit, 2023), explicitly teaching students what constitutes plagiarism and acceptable GenAI use, and sharing best practices among institutions (Xue et al., 2024). Future research could further investigate these ethical implications, as they are seldom discussed in depth in the current literature (notable exceptions include Katavic et al. (2023), and Santos and Cury (2023) in the reviewed corpus), and investigate how to mitigate such issues by design (e.g., privacy-by-design).

**Guideline 7 – Rethink targeted skills for CT:** *Evaluate which skills students should acquire within CT education. In particular, CT courses should balance instruction between algorithmic thinking and problem modeling, including finding patterns, designing abstraction, and problem decomposition.*

With students increasingly tempted to delegate simple coding exercises to a GenAI agent, the literature underlined the risk of over-reliance on outputs from GenAI and underscores the importance of critical thinking and self-regulation skills. The integration of GenAI into programming has lowered the barrier to entry for writing code, enabling a broader population to build software (Sarkar, 2023). Scholars have predicted a shift from *hard* coding to *vibe* coding - that is, directing GenAI agents that write code on one's behalf (Ray, 2025). While this shift might suggest that computational thinking skills are becoming obsolete, the need for people who can ensure the quality and security of software systems remains, making it important to reconsider the core skill set that future students should develop. Rather than translating ideas into syntax, learners will increasingly need to operate at higher levels of abstraction, such as designing software abstractions, reviewing and correcting AI-generated outputs, and ensuring system security, while engaging in core CT practices such as abstraction and iterative refinement through the evaluation and improvement of prompts and AI-generated responses (Hsu, 2025).

Learners must still critically assess and guide the output of GenAI tools for deeper learning, in contrast to shallow usage characterized by limited understanding of the content generated by GenAI (Shibani et al., 2024). Rather than focusing solely on algorithmic design, especially for non-computer science majors, CT curricula should also focus on problem formulation, i.e, the ability to articulate real-world challenges in structured ways that can be automated, whether through code or through well-crafted prompts in natural language.

## 5. Conclusion and Future Directions

In this systematic scoping literature review, we have mapped the landscape of GenAI-augmented learning experiences aimed at fostering computational thinking (CT) skills. A corpus of 38 peer-reviewed papers was assembled and analyzed to study four research questions. The results show a young, but rapidly growing research area, with mixed outcomes, some promising, others cautionary, which may partly reflect the variability in methodological rigor across studies. In particular, stronger designs such as randomized controlled trials remain limited, while many findings are derived from less controlled or exploratory settings. At the same time, the space is evolving rapidly beyond chat-based systems to other interaction paradigms and agentic workflows that may change what it means to teach CT with GenAI. A key future direction is to evaluate CT learning outcomes across a broader range of emerging models and levels of autonomy, and to periodically update evidence syntheses as GenAI ecosystem matures.

A recurring tension emerged between under-reliance and over-reliance on GenAI tools for CT. Beginners can easily delegate too much to GenAI when completing simple coding exercises, missing opportunities for meaningful learning. Conversely, more advanced students may fail to see the added value of GenAI in complex, open-ended projects where the tool cannot provide instant solutions. To address these challenges, we have inductively derived seven actionable design guidelines for educators and researchers. For instance, to support novices, GenAI should be positioned as a tutor that encourages critical thinking and active learning, rather than simply acting as a coder or debugger. For open-ended projects, we highlighted the importance

of helping students strengthen their computational thinking through strong problem modeling skills, i.e., problem decomposition, abstraction modeling, and pattern identification. These guidelines are to be treated as adaptive rather than fixed: As GenAI systems diversify, educators may need to recalibrate when and how to scaffold learner agency and responsibility to effectively reshape CT practices.

This review provided both a snapshot of current practices using GenAI in the curricula for CT and a foundation for future innovation in designing effective, ethical, and pedagogically sound GenAI-enhanced learning experiences for CT education. But it should be noted that, as a scoping review of an emerging field, this study does not aim to provide strong empirical or causal evidence, but rather to map and structure the existing body of work. Taking this context into consideration, and based on our synthesis, we note that the question for higher education is no longer how to integrate GenAI responsibly into existing learning objectives, but how to reconceive those objectives entirely, shifting from teaching students to produce computational solutions toward developing their capacity to formulate problems, reason and critically evaluate AI-generated outputs, and exercise the analytical judgment that remains distinctively human. Answering this question empirically is the central challenge for the next generation of research in this field.

## **6. Declaration of competing interest**

There is no potential conflict of interest in this study.

## 7. Data availability

Data will be made available upon request.

## References

- Abolnejadian, M., Alipour, S., Taeb, K., 2024. Leveraging chatgpt for adaptive learning through personalized prompt-based instruction: A cs1 education case study, in: CHI EA '24: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3613905.3637148>.
- Alshaikh, Z., Tamang, L., Rus, V., 2020. A socratic tutor for source code comprehension, in: Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21, Springer. pp. 15–19.
- Angeli, C., Voogt, J., Fluck, A., Webb, M., Cox, M., Malyn-Smith, J., Zagami, J., 2016. A k-6 computational thinking curriculum framework: Implications for teacher knowledge. *Journal of Educational Technology & Society* 19, 47–57.
- Ariza, Álvarez, J., Restrepo, M.B., Hernández, C.H., 2025. Generative ai in engineering and computing education: A scoping review of empirical studies and educational practices. *IEEE Access* .
- Baig, M.I., Yadegaridehkordi, E., 2024. Chatgpt in the higher education: A systematic literature review and research challenges. *International journal of educational research* 127, 102411.

- Batista, João, Mesquita, A., Carnaz, G., 2024. Generative ai and higher education: Trends, challenges, and future directions from a systematic literature review. *Information* 15, 676.
- Bryant, S., Romero, P., du Boulay, B., 2008. Pair programming and the mysterious role of the navigator. *International Journal of Human-Computer Studies* 66, 519–529.
- Cambaz, D., Zhang, X., 2024. Use of ai-driven code generation models in teaching and learning programming: a systematic literature review, in: *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, pp. 172–178.
- Chang, C.K., Chien, L.C.T., 2024. Enhancing academic performance with generative ai-based quiz platform, in: *2024 IEEE International Conference on Advanced Learning Technologies (ICALT)*, IEEE. URL: <https://ieeexplore.ieee.org/abstract/document/10645812>.
- Clarizia, F., Colace, F., Lombardi, M., Pascale, F., Santaniello, D., 2018. Chatbot: An education support system for student, in: *Cyberspace Safety and Security: 10th International Symposium, CSS 2018, Amalfi, Italy, October 29–31, 2018, Proceedings*. Springer. volume 10, pp. 291–302.
- Davis, F.D., et al., 1989. Technology acceptance model: Tam. Al-Suqri, MN, Al-Aufi, AS: *Information Seeking Behavior and Technology Adoption* 205, 5.
- Deng, R., Jiang, M., Yu, X., Lu, Y., Liu, S., 2024. Does chatgpt enhance

- student learning? a systematic review and meta-analysis of experimental studies. *Computers & Education* , 105224.
- Denny, P., Prather, J., Becker, B.A., Finnie-Ansley, J., Hellas, A., Leinonen, J., Luxton-Reilly, A., Reeves, B.N., Santos, E.A., Sarsa, S., 2024. Computing education in the era of generative ai. *Communications of the ACM* 67, 56–67.
- ElSayary, Areej, 2024. Integrating generative ai in active learning environments: Enhancing metacognition and technological skills. *Journal of Systems, Cybernetics and Informatics* 22, 34–37.
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., Gašević, D., 2025. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology* 56, 489–530. doi:<https://doi.org/10.1111/bjet.13544>.
- Frankford, E., Sauerwein, C., Bassner, P., Krusche, S., Brey, R., 2024. Aitutoring in software engineering education, in: *ICSE-SEET '24: Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training*, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3639474.3640061>.
- Følstad, A., Araujo, T., Law, E.L.C., Brandtzaeg, P.B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., et al., 2021. Future directions for chatbot research: An interdisciplinary research agenda. *Computing* 103, 2915–2942.

- Gabriella, A., Gui, A., Chanda, R.C., 2024. The use of chatbot and its impact on academic achievement, in: IEEE Symposium on Industrial Electronics & Applications (ISIEA), IEEE. URL: <https://ieeexplore.ieee.org/abstract/document/10607367>.
- Garcia, B, M., 2025. Teaching and learning computer programming using chatgpt: A rapid review of literature amid the rise of generative ai technologies. *Education and Information Technologies* , 1–25.
- Garg, A., Rajendran, R., 2024. The impact of structured prompt-driven generative ai on learning data analysis in engineering students, in: the 16th International Conference on Computer Supported Education (CSEDU 2024), Science and Technology publications. URL: <https://www.scitepress.org/Papers/2024/126930/126930.pdf>.
- Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., Sikdar, B., 2024. Privacy and security concerns in generative ai: a comprehensive survey. *IEEE Access* .
- Gottipati, S., Shim, K.J., Shankararaman, V., 2023. Ai for connectivism learning - undergraduate students' experiences of chatgpt in advanced programming courses, in: AMCIS, AIS. URL: [https://aisel.aisnet.org/amcis2023/sig\\_ed/sig\\_ed/16/](https://aisel.aisnet.org/amcis2023/sig_ed/sig_ed/16/).
- Groothuijsen, S., van den Beemt, A., Remmers, J.C., van Meeuwen, L.W., 2024. Ai chatbots in programming education: Students' use in a scientific computing course and consequences for learning. *Computers and Educa-*

- tion: Artificial Intelligence 7.0. URL: <https://www.sciencedirect.com/science/article/pii/S2666920X24000936>.
- Hsin, Wen-Jung, 2024. The effect of chatgpt: Student perspective and performance achievement. *Journal of Computing Sciences in Colleges* 39.0. URL: <https://dl.acm.org/doi/abs/10.5555/3665464.3665467>.
- Hsu, H.P., 2025. From programming to prompting: developing computational thinking through large language model-based generative artificial intelligence. *TechTrends* 69, 485–506.
- Hu, M., Assadi, T., Mahrooian, H., 2023. Explicitly introducing chatgpt into first-year programming practice: Challenges and impact, in: *International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, IEEE. URL: <https://ieeexplore.ieee.org/document/10398297>.
- Huesca, G., Martínez-Treviño, Y., Molina-Espinosa, J.M., Sanromán-Calleros, A.R., Martínez-Román, R., Cendejas-Castro, E.A., Bustos, R., 2024. Effectiveness of using chatgpt as a tool to strengthen benefits of the flipped learning strategy. *Education Sciences* 14.0. URL: <https://www.mdpi.com/2227-7102/14/6/660>.
- Hutson, James, 2024. Rethinking plagiarism in the era of generative ai. *Journal of Intelligent Communication* 3, 20–31.
- Jošt, G., Taneski, V., Karakatič, S., 2024. The impact of large language models on programming education and student learning outcomes. *Applied Sciences* 14.0. URL: <https://www.mdpi.com/2076-3417/14/10/4115>.

- Katavic, R., Pahuja, A., Syed, T.A., 2023. Navigating the use of chatgpt in classrooms: A study of student experiences, in: ICIS, AIS. URL: <https://aisel.aisnet.org/icis2023/learnandiscurricula/learnandiscurricula/15/>.
- Kazemitabaar, M., Chow, J., Ma, C.K.T., Ericson, B.J., Weintrop, D., Grossman, T., 2023. Studying the effect of ai code generators on supporting novice learners in introductory programming, in: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3544548.3580919>.
- Kazemitabaar, M., Hou, X., Henley, A., Ericson, B.J., Weintrop, D., Grossman, T., 2024. How novices use llm-based code generators to solve cs1 coding tasks in a self-paced learning environment, in: Koli Calling '23: Proceedings of the 23rd Koli Calling International Conference on Computing Education Research, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3631802.3631806>.
- Khatib, R., Mattalo, B., 2024. Enhancing learning experiences with genai chatbots: A tutorial approach, in: ICIS, AIS. URL: <https://aisel.aisnet.org/icis2024/learnandiscurricula/learnandiscurricula/17/>.
- Khemka, M., Houck, B., 2024. Toward effective ai support for developers: A survey of desires and concerns. *Communications of the ACM* 67, 42–49.
- Khotimah, K., Rusijono, A.M., 2024. Enhancing metacognitive and creativ-

- ity skills through ai-driven meta-learning strategies. *International Journal of Interactive Mobile Technologies* 18.
- Kosar, T., Ostojić, D., Liu, Y.D., Mernik, M., 2024. Computer science education in chatgpt era: Experiences from an experiment in a programming course for novice programmers. *Mathematics* 12.0. URL: <https://www.mdpi.com/2227-7390/12/5/629>.
- Kostopoulos, G., Gkamas, V., Rigou, M., Kotsiantis, S., 2025. Agentic ai in education: State of the art and future directions. *IEEE Access* 13, 177467–177491. doi:10.1109/ACCESS.2025.3620473.
- Law, K.M., Lee, V.C., Yu, Y.T., 2010. Learning motivation in e-learning facilitated computer programming courses. *Computers & Education* 55, 218–228.
- Lee, H.P.H., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., Wilson, N., 2025. The impact of generative ai on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers, in: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA. URL: <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3706598.3713778>, doi:10.1145/3706598.3713778.
- Li, C., Zhu, W., Xing, W., Guo, R., 2024. Analyzing student attention and acceptance of conversational ai for math learning: Insights from a randomized controlled trial, in: *LAK '24: Proceedings of the 14th Learning*

- Analytics and Knowledge Conference, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3636555.3636895>.
- Liao, J., Zhong, L., Zhe, L., Xu, H., Liu, M., Xie, T., 2024. Scaffolding computational thinking with chatgpt. *IEEE Transactions on Learning Technologies* 17.0. URL: <https://ieeexplore.ieee.org/abstract/document/10508087>.
- Lyu, W., Wang, Y., Chung, T., Sun, Y., Zhang, Y., 2024. Evaluating the effectiveness of llms in introductory computer science education: A semester-long field study, in: *L@S '24: Proceedings of the Eleventh ACM Conference on Learning @ Scale*, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3657604.3662036>.
- Macko, V., Ouazaki, A., Bartłomiejczyk, N., Monteiro, B., Felber, P., Holzer, A., 2025. Augmenting active learning with genai: Enhancement or impairment? evidence from a data visualization course, in: *Proceedings of the 20th European Conference on Technology Enhanced Learning (EC-TEL 2025)*. To appear.
- Margulieux, L.E., Prather, J., Reeves, B.N., Becker, B.A., Uzun, G.C., Loksa, D., Leinonen, J., Denny, P., 2024. Self-regulation, self-efficacy, and fear of failure interactions with how novices use llms to solve programming problems, in: *ITiCSE 2024: Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3649217.3653621>.
- Mezzaro, S., Gambi, A., Fraser, G., 2024. An empirical study on how large

- language models impact software testing learning, in: EASE '24: Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3661167.3661273>.
- Ohm, M., Bungartz, C., Boes, F., Meier, M., 2024. Assessing the impact of large language models on cybersecurity education: A study of chatgpt's influence on student performance, in: ARES '24: Proceedings of the 19th International Conference on Availability, Reliability and Security, IEEE. URL: <https://dl.acm.org/doi/abs/10.1145/3664476.3670446>.
- Okonkwo, C.W., Ade-Ibijola, A., 2021. Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence* 2, 100033.
- Ouaazki, A., Bergram, K., Farah, J.C., Gillet, D., Holzer, A., 2024. Generative ai-enabled conversational interaction to support self-directed learning experiences in transversal computational thinking, in: CUI '24: Proceedings of the 6th ACM Conference on Conversational User Interfaces, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3640794.3665542>.
- Ouaazki, A., Bergram, K., Holzer, A., 2023. Leveraging chatgpt to enhance computational thinking learning experiences, in: International Conference on Teaching, Assessment and Learning for Engineering (TALE), IEEE. URL: <https://ieeexplore.ieee.org/abstract/document/10398358>.
- Padiyath, A., Hou, X., Pang, A., Vargas, D.V., Gu, X., Nelson-Fromm, T., Wu, Z., Guzdial, M., Ericson, B., 2024. Insights from social shaping

- theory: The appropriation of large language models in an undergraduate programming course, in: ICER '24: Proceedings of the 2024 ACM Conference on International Computing Education Research, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3632620.3671098>.
- Peixoto, A.R., Glória, A., Silva, J.L., Pinto-Albuquerque, M., Brandão, T., Nunes, L., 2024. Use of programming aids in undergraduate courses, in: International Computer Programming Education Conference (ICPEC 2024), Schloss Dagstuhl. URL: <https://drops.dagstuhl.de/entities/document/10.4230/OASIcs.ICPEC.2024.20>.
- Peña-Fernández, S., Meso-Ayerdi, K., Larrondo-Ureta, A., Díaz-Noci, J., 2023. Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media. *El Profesional de la información* 32.
- Peyton, K., Unnikrishnan, S., 2023. A comparison of chatbot platforms with the state-of-the-art sentence bert for answering online student faqs. *Results in Engineering* 17, 100856.
- Qi, C., Tang, Y., Lei, Y., 2024. Does feedback from chatgpt help? investigating the effect of feedback from both teacher and chatgpt on students' learning outcomes, in: AMCIS, AIS. URL: [https://aisel.aisnet.org/amcis2024/is\\_education/is\\_education/29/](https://aisel.aisnet.org/amcis2024/is_education/is_education/29/).
- Qureshi, Basit, 2023. Chatgpt in computer science curriculum assessment: An analysis of its successes and shortcomings, in: ICSLT '23: Proceedings of the 2023 9th International Conference on e-Society, e-Learning and

- e-Technologies, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3613944.3613946>.
- Rajala, J., Hukkanen, J., Hartikainen, M., Niemelä, P., 2023. “Call me Kiran”–ChatGPT as a Tutoring Chatbot in a Computer Science Course, in: Mindtrek '23: Proceedings of the 26th International Academic Mindtrek Conference, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3616961.3616974>.
- Ranoliya, B.R., Raghuwanshi, N., Singh, S., 2017. Chatbot for university related faqs, in: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE. pp. 1525–1530.
- Rapp, A., Curti, L., Boldi, A., 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151, 102630.
- Ray, P.P., 2025. A review on vibe coding: Fundamentals, state-of-the-art, challenges and future directions. *Authorea Preprints* .
- Rich, P.J., Egan, G., Ellsworth, J., 2019. A framework for decomposition in computational thinking, in: Proceedings of the 2019 ACM conference on innovation and technology in computer science education, pp. 416–421.
- Samarakoon, P., Asanka, D., Jayalal, S., Jayalath, N., 2024. Analyzing the learning effectiveness of generative ai for software development for undergraduates in sri lanka, in: International Research Conference

- on Smart Computing and Systems Engineering (SCSE), IEEE. URL: <https://ieeexplore.ieee.org/abstract/document/10550837>.
- Santos, O.L.D., Cury, D., 2023. Challenging the confirmation bias: Using chatgpt as a virtual peer for peer instruction in computer programming education, in: 2023 IEEE Frontiers in Education Conference (FIE), IEEE. URL: <https://ieeexplore.ieee.org/abstract/document/10343247/>.
- Sarkar, A., 2023. Will code remain a relevant user interface for end-user programming with generative ai models?, in: Proceedings of the 2023 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software, pp. 153–167.
- Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A., Leimeister, J.M., 2024. Charting the evolution and future of conversational agents: A research agenda along five waves and new frontiers. *Information Systems Frontiers* 26, 729–754.
- Sengewald, J., Wilz, M., Lackes, R., 2024. Ai-assisted learning feedback: Should gen-ai feedback be restricted to improve learning success? a pilot study in a sql lecture, in: ECIS, AIS. URL: [https://aisel.aisnet.org/ecis2024/track13\\_learning\\_teach/track13\\_learning\\_teach/12/](https://aisel.aisnet.org/ecis2024/track13_learning_teach/track13_learning_teach/12/).
- Sethi, Farhana, 2020. Faq (frequently asked questions) chatbot for conversation. *International Journal of Computer Sciences and Engineering* 8, 7–10.
- Sezer, B., Yilmaz, R., 2019. Learning management system acceptance scale

- (lmsas): A validity and reliability study. *Australasian Journal of Educational Technology* 35.
- Sheese, B., Liffiton, M., Savelka, J., Denny, P., 2024. Patterns of student help-seeking when using a large language model-powered programming assistant, in: *ACE '24: Proceedings of the 26th Australasian Computing Education Conference*, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3636243.3636249>.
- Shibani, A., Knight, S., Kitto, K., Karunanayake, A., Buckingham Shum, S., 2024. Untangling critical interaction with ai in students' written assessment, in: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA. URL: <https://dl.acm.org/doi/10.1145/3613905.3651083>, doi:10.1145/3613905.3651083.
- Sinha, S., Basak, S., Dey, Y., Mondal, A., 2020. An educational chatbot for answering queries, in: *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*. Springer, pp. 55–60.
- Solyst, J., Yang, E., Xie, S., Hammer, J., Ogan, A., Eslami, M., 2024. Children's overtrust and shifting perspectives of generative ai. arXiv preprint arXiv:2404.14511 .
- Stack Overflow, 2023. Stack overflow developer survey 2023. URL: <https://survey.stackoverflow.co/2023/#section-sentiment-and-usage-ai-tools-in-the-development-process>. [Online; accessed 18-March-2025].

- Su, J., Yang, W., 2023. Unlocking the power of chatgpt: A framework for applying generative ai in education. *ECNU Review of Education* 6, 355–366.
- Sun, Y., Jang, E., Ma, F., Wang, T., 2024. Generative ai in the wild: Prospects, challenges, and strategies, in: *CHI'24*, pp. 1–16.
- Takkouche, B., Norman, G., 2011. Prisma statement. *Epidemiology* 22, 128.
- Tsai, M.J., Liang, J.C., Hsu, C.Y., 2021. The computational thinking scale for computer literacy education. *Journal of Educational Computing Research* 59, 579–602.
- Vadaparty, A., Zingaro, D., Iv, D.H.S., Padala, M., Alvarado, C., Benario, J.G., Porter, L., 2024. Cs1-llm: Integrating llms into cs1 instruction, in: *ITiCSE 2024: Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3649217.3653584>.
- Wang, Z., Chen, L., Anderson, T., 2014. A framework for interaction and cognitive engagement in connectivist learning contexts. *International Review of Research in Open and Distributed Learning* 15, 121–141.
- Warr, M., Oster, N.J., Isaac, R., 2024. Implicit bias in large language models: Experimental proof and implications for education. *Journal of Research on Technology in Education* , 1–24.
- Webster, J., Watson, R.T., . Analyzing the past to prepare for the future: writing a literature review. *MIS Q* 26.

- Weidlich, J., Gašević, D., 2025. Chatgpt in education: An effect in search of a cause .
- Welsh, Matt, 2022. The end of programming. *Communications of the ACM* 66, 34–35.
- Wing, Jeannette, 2011. Research notebook: Computational thinking—what and why. *The Link Magazine* 6, 20–23.
- Wing, M, J., 2006. Computational thinking. *Communications of the ACM* 49, 33–35.
- Xue, Y., Chen, H., Bai, G.R., Tairas, R., Huang, Y., 2024. Does chatgpt help with introductory programming?an experiment of students using chatgpt in cs1, in: *ICSE-SEET '24: Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training*, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3639474.3640076>.
- Yang, A.C.M., Lin, J.Y., Lin, C.Y., Ogata, H., 2024a. Enhancing python learning with pytutor: Efficacy of a chatgpt-based intelligent tutoring system in programming education. *Computers and Education: Artificial Intelligence* 7.0. URL: <https://www.sciencedirect.com/science/article/pii/S2666920X24001127>.
- Yang, S., Zhao, H., Xu, Y., Brennan, K., Schneider, B., 2024b. Debugging with an ai tutor: Investigating novice help-seeking behaviors and perceived learning, in: *ICER '24: Proceedings of the 2024 ACM Con-*

ference on International Computing Education Research, ACM. URL: <https://dl.acm.org/doi/abs/10.1145/3632620.3671092>.

Yilmaz, R., Yilmaz, F.G.K., 2023. The effect of generative artificial intelligence (ai)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence 4.0*. URL: <https://www.sciencedirect.com/science/article/pii/S2666920X23000267>.

**Declaration of Interest Statement**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: