

Primary AI Literacy Pedagogy before and after Generative AI: A Critical Review

Akorfa Aku Nyatefe ^{*1}, Isaac Mawuena Agbogah ², Emmanuel Akwandoh ³,
Emmanuel Yeboah ⁴, Enock Peprah ³, Sitsofe Akosua Nyatefe ⁵, and Madeakpe
Gertrude Nyatefe ⁶

¹ School of Teacher Education, Nanjing University of Information Science and Technology, Nanjing, 210044, China

² School of Business, Nanjing University of Information Science and Technology, Nanjing, 210044, China

³ School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, 210044, China

⁴ Faculty of Geographical Science and Engineering, Henan University, Kaifeng, Henan Province, China

⁵ The New Jerusalem School, Accra, Ghana

⁶ School of Graduate Studies, University of Education, Winneba, Ghana

nyatefeakorfa@gmail.com

Received: August 19, 2026; **Accepted:** October 20, 2026; **Published:** October 31, 2026

Abstract

Two recent reviews of AI literacy in primary education closed their searches in 2024, before most classroom research on generative AI had appeared. This critical review narrows that gap. It takes the pre-2023 literature from those reviews and adds studies published through October 2026 in which children roughly aged six to twelve used generative AI tools such as ChatGPT. Twenty empirical papers are examined: eight from before 2023, two on children's conceptions of AI, and ten on generative AI (nine independent samples). The synthesis covers pedagogical approaches, learning outcomes, assessment, and challenges. Earlier studies mostly taught children about AI, while the generative-AI studies mostly test learning with AI, so the two literatures are not directly comparable. A continuity-with-adaptation framework is offered as a hypothesis, not a finding: hands-on, teacher-guided work is the common thread, and generative AI adds the need to judge fluent but unreliable output and to attend to privacy and age limits. Controlled studies have appeared only recently, cluster in a few settings, and none of those identified report a delayed follow-up. Early childhood was found to be underexplored. The review closes with a research agenda and cautious suggestions for teachers.

Keywords: AI literacy, generative AI, primary education, ChatGPT, pedagogy

1. Introduction

Children now meet AI in recommendation feeds and voice assistants, and since late 2022, in tools that write, draw, and answer questions on request. Defining, teaching, and assessing AI literacy for children aged about six to twelve is not a new question. Generative AI has changed what teachers and researchers must address.

Two reviews anchor the field. Yim and Su (2025) searched Scopus and Web of Science and selected 25 empirical studies on AI literacy education in primary schools. Aravantinos et al. (2024) reviewed 35 Scopus-indexed articles on AI in primary education for children aged four to twelve, following PRISMA guidelines.

Neither review could cover most of the work on generative AI. Yim and Su's search closed in March 2024, and Aravantinos et al. submitted their article in June 2024. Neither was designed around generative AI tools, and ChatGPT-type tools only began to appear in classroom studies from 2023.

A review of generative AI in K-12 education points to the same gap from the other side. Alfarwan (2025) analysed 30 empirical studies and found that most focused on high school students and pre-service teachers. Primary, elementary, and secondary levels each accounted for only about two studies, and none addressed early childhood education.

1.1. Why Primary Education Needs Its Own Review

Primary-aged children differ from secondary students in ways that bear on how AI literacy is taught and studied. Piaget and Inhelder (1969) describe the school years from roughly seven to eleven as a period of concrete operational thinking, in which reasoning is still tied to objects and situations a child can handle or see. Judging whether an AI output is accurate, biased, or invented asks for more abstract checking than that. Activities designed for teenagers, such as spotting a chatbot's errors against a reference text, may need to be rebuilt around comparisons children can actually carry out.

Children also arrive with ideas of their own. In a Finnish study of fifth and sixth graders, Mertala and Fagerlund (2024) identified three kinds of misconception: AI as people's cognitive processes, AI as a human-like entity, and AI as a machine with pre-installed knowledge. Most of the children rated their own AI knowledge as low.

The setting differs as well. In many systems, primary teachers are generalists who also teach reading and mathematics, so AI literacy competes with foundational subjects for time and for training.

1.2. Aim and Research Questions

This review aims to map what is known about teaching and learning AI literacy with children aged about six to twelve before and after generative AI, and to ask how far the two bodies of work can be read together. It is guided by four questions:

- RQ1: Which pedagogical approaches were used to teach AI literacy to children aged about six to twelve before and after generative AI?
- RQ2: Which learning outcomes were reported?
- RQ3: How were these outcomes assessed?
- RQ4: Which challenges remain, and how do the two periods relate?

2. Literature Review and Positioning

The two anchor reviews describe a field built mainly on hands-on work. Yim and Su (2025) report that constructionism and constructivist frameworks dominated study designs, that project-based learning, programming, and human-agent interaction were the common methods, that mixed-methods designs were most frequent, and that studies reported positive academic, affective, and behavioral outcomes. Aravantinos et al. (2024) sort the pedagogy of AI tools into constructivism, experiential learning, AI-assisted learning, and project-based learning.

Other work supplies frameworks and wider context. The AI4K12 initiative set out what every child should know about AI (Touretzky et al., 2019). Ng et al. (2021) reviewed how AI literacy has been conceptualised. Casal-Otero et al. (2023) reviewed AI literacy across K-12. Yim (2024) analysed 17 frameworks in a scoping review and proposed one for young students that acknowledges both human and non-human agents. None of these was built around generative AI in primary classrooms.

Figure 1 gives the key dates. Yim and Su's search closed in March 2024, Aravantinos et al. submitted in June 2024, and Alfarwan's search was run on 19 February 2025. All three therefore fall short of the generative-AI studies that appeared through 2025 and 2026, which is where this review adds evidence.

A caution applies to everything that follows. Studies from before 2023 mostly taught children about AI: they built, trained, or tested something. The generative-AI studies mostly examine learning with AI in a school subject such as reading or writing. Because the two groups measure different things, this review compares them descriptively and does not claim that one period's results carry over to the other.

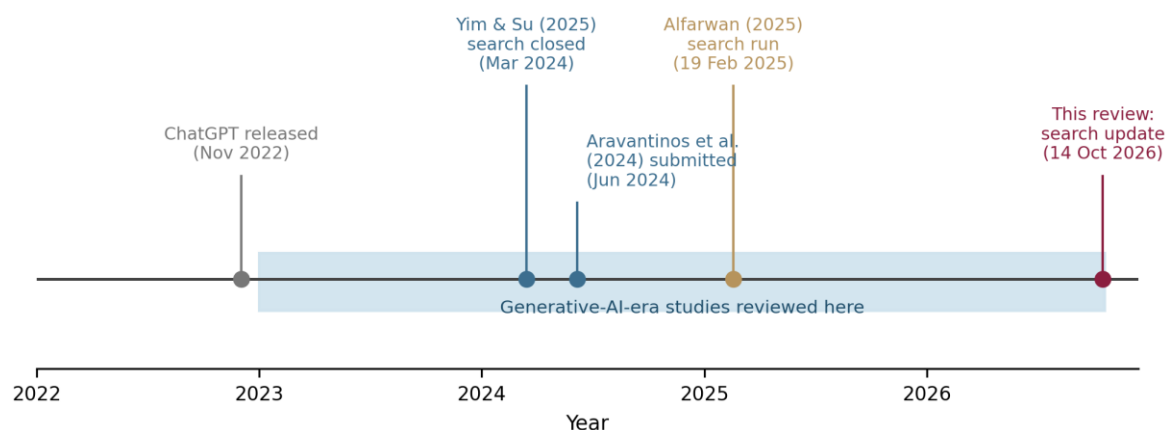


Figure 1. Key Dates for the Three Reviews this Paper Builds on, and the Search Update for this Review

3. Methodology

3.1. Type of Review

This is a critical review. It is not a systematic review, and it does not claim to meet PRISMA-ScR reporting standards. The screening and synthesis were carried out by the author team, without an independent second screener, and without institutional access to Scopus, Web of Science, IEEE Xplore, or ACM Digital Library. For a database-verified account of the pre-2024 literature, readers should use Yim and Su (2025) and Aravantinos et al. (2024).

3.2. Sources and Eligibility

Table 1 sets out the source groups, the eligibility criteria, and how each group was checked. Pre-2023 studies were identified through Yim and Su (2025). For this revision, abstracts were retrieved for five of the eight (Druga & Ko, 2021; Melsión et al., 2021; Shamir & Levin, 2021, 2022; Voulgari et al., 2021). Abstracts were not retrieved for the other three (Ng et al., 2022; Rodríguez-García et al., 2021; Vartiainen et al., 2020), so no findings are attributed to them. Studies from 2023 onward were retrieved directly. Abstracts or equivalent descriptions were retrieved in every case, and the full texts of Balantekin (2026) and Kızıldaş (2026) were read. The original text of Han et al. (2024) was not accessed, and its sample sizes come from a later paper's description of the study.

3.3. Search and Screening

The update for the generative-AI period was run on 14 October 2026 with a general web search tool. Queries paired generative AI terms (“generative AI,” “ChatGPT”) with primary school terms (“primary school,” “elementary,” “fourth grade”), and publisher, repository, and indexing pages reached through the results were used to confirm details. The first ten results for each

query were screened on title and abstract against Table 1. In addition, the Appendix lists the queries. Earlier exploratory searches for the first draft were not logged, and record counts beyond ten results per query were not kept.

Table 1. Source groups, eligibility criteria, and verification

Source group	Period	Eligibility	How checked
Pre-2023 empirical studies	2020–2022	Empirical. Participants in primary grades (about ages 6–12), or mixed-age samples that include them, flagged as boundary cases. AI literacy teaching, tools, or outcomes.	Identified through Yim & Su (2025). Abstracts retrieved for five of eight; no findings reported for the other three.
Children’s conceptions of AI	2024	Empirical. Primary-aged children’s ideas about AI or machine learning. No generative AI tool.	Abstracts retrieved. The machine-learning study is an arXiv preprint.
Generative-AI empirical studies	2023–2026	Empirical. Primary-grade participants, or mixed-age samples that include them. A generative AI tool was used or examined with children.	Abstracts retrieved for all; full text read for two (Balantekin, 2026; Kızıldaş, 2026).
Context	2019–2026	Reviews, frameworks, and policy documents.	Retrieved directly.

Note. Boundary cases because of participant age: Melsión et al. (2021), ages 10–14; Rodríguez-García et al. (2021), ages 10–16; Voulgari et al. (2021), primary and secondary students; Vartiainen et al. (2023), K–9; Jauhainen and Garagorry Guerra (2023, 2024), ages 8–14 in grades 4–6.

Two papers reporting the same pupils were counted as one sample (Jauhainen & Garagorry Guerra, 2023, 2024). The generative-AI literature therefore comprises ten papers and nine independent samples. Figure 2 shows the included empirical papers by year and period.

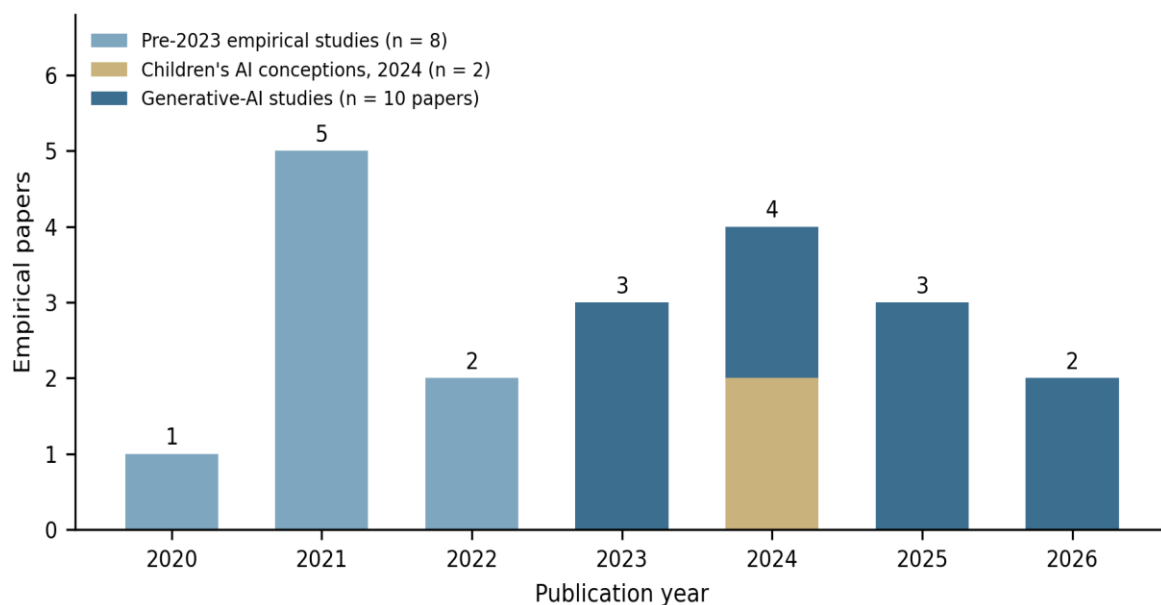


Figure 2. Empirical Papers Included in this Review, by Publication Year and Period (N = 20)

4. Thematic Synthesis

4.1. Before 2023: Teaching Children About AI (RQ1)

Most pre-2023 studies had children build, train, or test an AI system. Shamir and Levin (2021) gave a small group of 12-year-old elementary students a programmable environment for constructing neural networks. They reported high engagement and that the environment helped make machine learning understandable. Shamir and Levin (2022) went on to compare course formats in elementary school, including a learning-by-teaching course, and the groups improved in different aspects of computational thinking.

Druga and Ko (2021) ran a four-week after-school study with 52 children aged 7 to 12 who trained and coded their own smart programs. The children became more skeptical of what smart devices can do, and they shifted their attribution of agency from the devices to the people who program them. Melsión et al. (2021) tested an explainability technique with 78 children aged 10 to 14 and found that they better recognised biased predictions after working with the interpretable model. Voulgari et al. (2021) described the design rationale for a game, ArtBot, together with initial feedback from students and teachers across primary and secondary education.

Three further studies in the Yim and Su sample fit the same pattern by their titles: an online intervention using the LearningML tool with students aged 10–16 (Rodríguez-García et al., 2021), digital story writing to develop AI literacy among primary students (Ng et al., 2022), and machine learning with very young children (Vartiainen et al., 2020). No findings from these three are reported here.

In these studies, children acted on a system they could inspect and change. That is an observation about the studies, not a tested explanation of why they worked.

4.2. After 2022: Generative AI in Primary Classrooms (RQ1)

Ten papers report generative AI work with primary-aged children, from nine independent samples. Table 2 summarises each. The design-strength column is the authors' judgement from the reported methods and is explained as a note under the table.

4.2.1. Early classroom trials

Jauhiainen and Garagorry Guerra (2023, 2024) tested ChatGPT-3.5 and 4 with 110 pupils aged 8 to 14 in grades 4 to 6 of two Uruguayan schools. The tools generated and adjusted lesson text, illustrations, and exercises to pupils' knowledge levels during the lesson. The authors report

that this raised motivation and performance. The abstracts describe no comparison group, so the result cannot be separated from the novelty of the lesson.

Table 2. Generative-AI Studies with Primary-Aged Children (2023–2026)

Study	Country	Sample	Design	Key result, as reported	Design strength	Main limitation
Jauhiainen & Garagorry Guerra (2023, 2024)	Uruguay	110 pupils, grades 4–6 (ages 8–14), two schools	Classroom test lessons; no comparison group described	Better motivation and performance with personalised material	Low	One lesson reported in two papers; no delayed measure described
Relmasira et al. (2023)	Indonesia	77 reflection papers; fifth grade	Design-based research; three-session intervention	Design principles for AI literacy	Descriptive	No comparison; reflections as data
Vartiainen et al. (2023)	Finland	Not reported in abstract	Exploratory qualitative study, K–9	Relations with generative AI shown in digital artefacts and peer discussion.	Descriptive	Mixed K–9 ages
Han et al. (2024)	USA	12 parent-child pairs (ages 8–12); 16 educators	Workshops and interviews	Perspectives on generative AI in elementary literacy; no learning outcomes	Descriptive	Perceptions only
Song et al. (2025)	USA	79 fifth graders; 6 teachers	Qualitative: interviews, open-ended surveys, observation	Six themes, 19 subthemes on perceived benefits and challenges	Descriptive	Perceptions only; one technology
Kızıldaş (2025)	Türkiye (Van)	62 (31 + 31), age 10	Quasi-experimental, pre/post, control group	Higher creative-writing and self-efficacy scores with ChatGPT feedback	Moderate	Short-term; scale-based outcomes
Yang et al. (2025)	China	52 fifth graders	Quasi-experimental usability study of ChatGPT-MWPS	Usability findings, not achievement gains	Low	One class; one system
Balantekin (2026)	Türkiye (Bursa)	106 reported; four classes	Solomon four-group, class-level random assignment; 20 interviews	$\eta^2=.29$ (reading errors), .40 (comprehension), .16 (motivation)	Higher, with caveats	Four intact classes; researcher administered and scored, no reliability statistic; 10 weeks
Kızıldaş (2026)	Türkiye (Van)	68 (34 + 34), age 10	Quasi-experimental, nonequivalent control, ANCOVA	$\eta^2=.33$ (summarisation skill), .48 (attitude)	Moderate	One class per condition; six weeks; no follow-up

Note. Design strength: Higher = random assignment with pre- and post-measures; Moderate = control group without random assignment; Low = no comparison group, or a usability focus; Descriptive = qualitative or design-based work not built to estimate effects. η^2 values are the authors' own. Sample sizes for Han et al. (2024) are taken from a later paper's description of the study.

4.2.2. Perception and design studies

Four studies describe how children and adults experienced generative AI. Relmasira et al. (2023) ran a three-session intervention in an Indonesian elementary school and analysed 77

student reflection papers through design-based research to produce design principles. Vartiainen et al. (2023) examined children co-creating digital art with generative AI across K–9 and report that relationships with the technology showed up in the artefacts and in peer discussion. Han et al. (2024) gathered the views of parents, children, and educators on generative AI in elementary literacy teaching. Song et al. (2025) analysed interviews, open-ended surveys, and observations with 79 fifth graders and six teachers using a generative AI tool for creative mathematical writing, and identified six themes on perceived benefits and challenges. These studies help with design, but none was built to measure learning.

4.2.3. Comparative studies

Four studies compared conditions or measured outcomes. Kızıldaş (2025) used a quasi-experimental pre/post design with 31 children per group in Van, Türkiye, in which ChatGPT supplied feedback on writing under teacher guidance. The study reports higher creative-writing and writing self-efficacy scores in the ChatGPT group. Yang et al. (2025) built ChatGPT-MWPS, a tutoring system for mathematical word problems, and evaluated it with 52 fifth graders in China. It is a usability study, so it says little about mathematics achievement.

Balantekin (2026) is the most rigorous design so far. Four fourth-grade classes in two schools in Bursa, Türkiye, were assigned at the class level to a Solomon four-group design. Students wrote their own prompts; ChatGPT-4 generated texts, questions, and visuals, and the printed materials were used in ordinary lessons over ten weeks. Compared with control classes, the experimental classes showed fewer reading errors ($\eta^2 = .289$), higher comprehension ($\eta^2 = .399$), and higher reading motivation ($\eta^2 = .16$). Twenty students were also interviewed. The caveats are real. There were only four intact classes. The researcher administered the instruments and did the scoring and interview coding, and no inter-rater statistics are reported. The intervention lasted ten weeks in one cultural and linguistic setting.

Kızıldaş (2026) compared 34 fourth graders using ChatGPT with 34 controls in one school in Van, Türkiye. In the experimental class, the whole group used a single terminal on an interactive whiteboard with the teacher's ChatGPT Plus account. Students wrote summaries, took turns comparing them with ChatGPT's feedback, and the teacher reviewed and confirmed that feedback. Post-test summaries were written without ChatGPT. After adjusting for pre-test scores, summarisation skill ($\eta^2 = .334$) and attitude toward summarising ($\eta^2 = .481$) were higher in the experimental class. There was one class per condition, so teacher and class are tangled up with the treatment. The intervention lasted six weeks. A second researcher independently scored the summaries (Cohen's $\kappa = .91$).

4.2.4. How much weight the effect sizes can bear

The effect sizes in these two papers look large. They come from single-site studies of six to ten weeks, with one or two classes per condition, measured immediately after the intervention. The authors' own recommendations that ChatGPT be integrated into classrooms go further than that evidence. The results are promising and consistent in direction, but they are not settled.

Across the ten papers, the tool is mostly used for personalisation, feedback, or co-creation, and an adult controls its use. In the two papers read in full, adults supervised use of the tool, and in Kızıldaş (2026) it ran through the teacher's account on a shared whiteboard. Three of the studies (Yang et al., 2025; Balantekin, 2026; Kızıldaş, 2026) also ground their interventions in learning theory. Three studies cannot establish a trend, but it would be worth watching whether later studies do the same.

4.3. Reported Outcomes (RQ2)

Reported outcomes fall into four groups, shown in Table 3: conceptual understanding of AI, academic performance with AI, affective outcomes, and perceptions or usability.

Gains in the generative-AI studies are in academic performance and affect. Gains in the earlier studies are mostly in understanding. That difference is the central problem for any comparison, and Section 5 returns to it.

Misconceptions are the one concern that appears in both periods. Mertala et al. (2024), a preprint, asked 197 Finnish primary schoolers what machine learning is without first telling them that humans teach machines. The answers were varied, and a substantial share described it as learning via or about machines. That study was published in 2024 but concerns machine learning in general, not generative AI, so it is grouped with the earlier literature by topic. It is a preprint, and no peer-reviewed version was found. The related peer-reviewed study of AI misconceptions (Mertala & Fagerlund, 2024) is cited in Section 1.1.

Nearly all measurements in the generative-AI studies were taken straight after the intervention. None of the ten papers identified reports a delayed follow-up, so novelty effects remain a plausible explanation for part of the gains.

4.4. Assessment (RQ3)

Yim and Su (2025) report that mixed methods were most common among the studies they reviewed, using surveys, interviews, and artefacts. Alfarwan (2025) found that quasi-experimental designs and psychological measures were most common in generative-AI studies, and notes calls in that literature for more objective measures.

Table 3. Reported outcomes by study and period

Group	Outcome	Studies	Period
Conceptual understanding of AI	Machine learning made understandable (small group)	Shamir & Levin (2021)	Pre-2023
	Different aspects of computational thinking improved, by course format	Shamir & Levin (2022)	Pre-2023
	Attribution of agency shifted from devices to programmers; more skepticism	Druga & Ko (2021)	Pre-2023
	Better recognition of biased predictions (N = 78, ages 10–14)	Melsión et al. (2021)	Pre-2023
	Varied and often mistaken ideas about AI and machine learning	Mertala & Fagerlund (2024); Mertala et al. (2024, preprint)	2024, no generative AI
Academic performance with AI	Creative writing	Kızıldaş (2025)	Generative AI
	Reading errors and comprehension	Balantekin (2026)	Generative AI
	Summarisation skill	Kızıldaş (2026)	Generative AI
	Motivation and performance with personalised lessons	Jauhainen & Garagorry Guerra (2023, 2024)	Generative AI
Affective	Engagement	Shamir & Levin (2021)	Pre-2023
	Writing self-efficacy	Kızıldaş (2025)	Generative AI
	Reading motivation	Balantekin (2026)	Generative AI
	Attitude toward summarising	Kızıldaş (2026)	Generative AI
Perceptions and usability	Perceived benefits and challenges; stakeholder views	Song et al. (2025); Han et al. (2024)	Generative AI
	System usability	Yang et al. (2025)	Generative AI
	Initial feedback on a game-based tool	Voulgari et al. (2021)	Pre-2023

Balantekin (2026) and Kızıldaş (2026) move in that direction by using inventories or scoring rubrics for reading and summarising alongside attitude scales. They differ in how the scoring was checked. In Kızıldaş (2026), a second researcher scored the summaries independently, with $\kappa = .91$. In Balantekin (2026), the researcher administered the instruments and did the scoring and interview coding, and no inter-rater statistic is reported.

Two constructs are easy to blur: whether a child understands how an AI system works and can judge its output (AI literacy), and whether a child performed better in a subject with AI support (learning with AI). The generative-AI studies measured the second. Each was designed for that purpose, so this is not a flaw. It does leave an open question: because no study measured both in the same children, we do not know whether the academic gains come with, are independent of, or come at the cost of AI literacy.

Several adaptations of earlier methods would help. Artefact-scoring rubrics of the kind used with neural-network construction could become checklists for judging whether a generated answer is plausible and consistent with a trusted source. Pre/post probes of how children think ChatGPT knows things could follow the conceptual-change approach of Mertala et al. (2024). Observation protocols could record what teachers do when a tool produces an error.

4.5. Remaining Challenges (RQ4)

4.5.1. Teacher preparation

Alfarwan (2025) reports that about 90% of the reviewed generative-AI studies emphasised teachers' professional development. The generative-AI studies read in full here also rely on teachers or researchers to run the tool, which assumes training.

4.5.2. Reliability of output

Generative tools produce fluent text that is sometimes wrong. In Balantekin's (2026) pilot, students first tested ChatGPT with questions whose answers they already knew, such as the capital of their country. They then read the generated texts critically against their own prompts, and some students identified limitations in the generated output and suggested improvements. These are activities, not measured outcomes. None of the studies read for this review measured whether children's skill at judging AI output improved or lasted once support was removed.

4.5.3. Privacy, consent, and age limits

OpenAI's terms of use (2026) require users to be at least 13, or the minimum age in their country, and to have a parent's or guardian's permission if under 18. Primary-aged children are below that threshold. The terms also say that OpenAI may use content to develop and improve its services. These conditions matter for any classroom use of a general chatbot.

The two papers read in full show how researchers handled this. Balantekin (2026) reports parental and school consent, a secure connection supplied by the education ministry, and researcher supervision. In Kızıldaş (2026), the tool was used through a teacher's account on a shared whiteboard, with parental consent collected through teachers. The other abstracts do not say how privacy was handled. Studies should report this, and schools should not enter children's personal information into general-purpose chatbots.

4.5.4. Geographic concentration

The nine generative-AI samples come from six countries: Türkiye (three), the United States (two), and one each from Uruguay, Indonesia, Finland, and China. Yim and Su (2025) highlight

China, the United States, Finland, and Israel, and Alfarwan (2025) highlights Taiwan, China, the United States, and Germany. Türkiye is not among the countries those reviews highlight, so the Turkish studies are a real, if modest, broadening. Two of the three are by the same author in the same city (Van), so they are narrower than the count suggests.

Uruguay is the only Latin American setting, and no study from Africa, South Asia, or Southeast Asia apart from Indonesia was found. Alfarwan (2025) likewise identified no African studies in the databases searched. Whether this is an absence of research or a limit of the databases and search languages used cannot be settled here. Paid access to tools, English-language publishing, and national policy priorities may all contribute. The practical risk is that guidance rests on evidence from a few well-resourced settings.

4.5.5. *Early childhood*

Early childhood was found to be underexplored, and Alfarwan (2025) found no studies either. Young children may encounter conversational AI at home before any school introduces it. Research with this age group would require methods for capturing ideas from children who cannot yet express themselves fully in words.

5. Discussion

5.1. Can the Two Literatures Be Read Together?

Only with care. Table 4 compares the periods on the same points. The comparison shows that the studies differ in aim, tool, and outcome, so differences in findings may reflect what was measured as much as what changed.

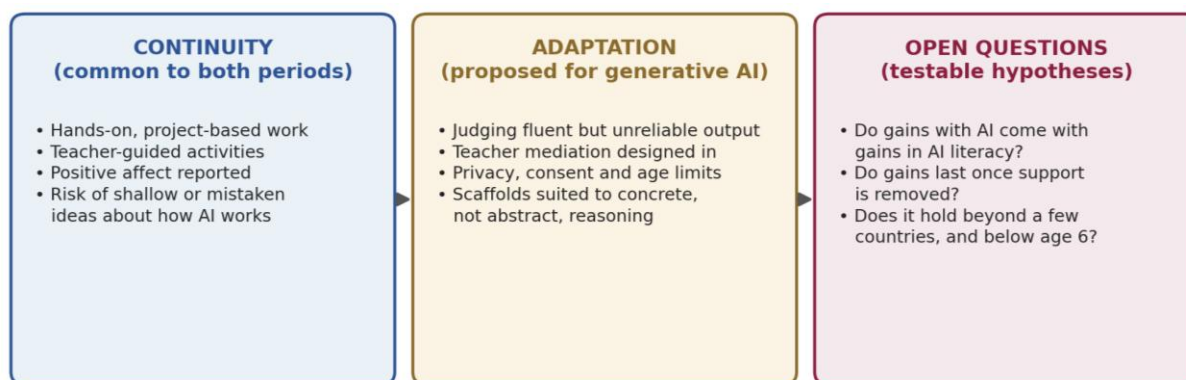
Table 4. *The Two Periods Compared on the Same Points*

Point	Before 2023	Generative AI (2023–2026)
Main aim	Teach children about AI: how it learns, can be biased, is built	Use AI to support learning in a subject, or study how children respond to it
Typical tool	Systems children can train, program, or inspect	Conversational or image tools whose workings children cannot inspect
Who controls the tool	Children, in hands-on activities	Teacher or researcher, in the two papers read in full
Typical outcome	Understanding, engagement, recognising bias	Subject performance, self-efficacy, motivation, attitude
Typical assessment	Small samples; observation; tests of understanding	Pre/post scales and rubrics; qualitative perceptions
Evidence base here	Eight papers, five read at abstract level, plus Yim & Su (2025)	Nine samples, mostly read at abstract level, two in full

The evidence is also unbalanced. Yim and Su reviewed 25 studies, many more than the nine generative-AI samples, and the older work is characterised largely through that review while the newer work was read more directly. The newer effect sizes come from short, single-site designs. Both facts argue against reading the periods as a clean before-and-after.

5.2. Continuity with Adaptation: A Proposal

One reading of the two bodies of work is that they share a core and differ at the edges. Figure 3 sets this out as a proposal. The shared core is hands-on, teacher-guided activity and positive affect. A shared risk is shallow or mistaken ideas about how AI works. The possible adaptations are specific to generative AI: judging fluent but unreliable output, designing teacher mediation in, attending to privacy and age limits, and choosing scaffolds that suit concrete reasoning.



A proposal drawn from two bodies of work that measure different things (learning about AI, learning with AI). It is a hypothesis to test, not a finding.

Figure 3. The Continuity-With-Adaptation Proposal. It is a Hypothesis Drawn from two Literatures that Measure Different Things

This is an argument, not a finding. For the latter period, no study identified here compares a hands-on, project-based approach with any other, so “continuity” describes a shared pattern and not evidence that the approach works. The proposal does lead to testable questions, listed in the right-hand box of Figure 3. Frameworks such as those of Touretzky et al. (2019) and Ng et al. (2021) already treat evaluation as a distinct competency, and Yim’s (2024) framework for young students could be examined for how far it covers evaluating generative output. This review cannot say whether it does.

6. Conclusion

6.1. Remaining Gaps

The generative-AI literature on primary education is still small, and controlled studies have appeared only in the last two years. Four gaps stand out. No study measured AI literacy and

subject learning in the same children. None of the studies identified report a delayed follow-up. The evidence comes from a few countries. Early childhood was found to be underexplored.

6.2. Limitations

This is a critical review without access to the major databases, so it may have missed relevant studies, especially recent ones and those published outside English. Pre-2023 studies are characterised mainly through Yim and Su (2025) and, for five of eight, through abstracts. Most generative-AI studies were read at the abstract level, so details such as sample sizes and limitations may be incomplete. One cited source on children's conceptions of machine learning is a preprint. Search logs for earlier exploratory searches were not kept. The continuity-with-adaptation proposal and the design-strength ratings are the authors' judgement.

6.3. Recommendations for Research

- Measure AI literacy and subject learning together in the same sample.
- Add delayed follow-up measures, and test whether children can still judge AI output once support is removed.
- Report scoring procedures and reliability, randomise or counterbalance at a level above one class per condition, and report how privacy and consent were handled.
- Study settings beyond the few countries now represented, and children younger than six.

6.4. Practical Suggestions for Teachers

The evidence is early, from few settings, and mostly without follow-up, so these are suggestions, not firm advice.

- Check the tool's age and consent terms first. Run it yourself, on a shared display, and keep children's personal information out of prompts.
- Give children chances to check outputs against things they already know or can look up, as in the pilot activity in Balantekin (2026).
- Keep the teacher in the loop. In the two papers read in full, the teacher or researcher ran the tool and reviewed its output.
- Build on hands-on activities already in use, and treat any reported gains as preliminary.
- Plan for training time. Teacher preparation is the most commonly reported need in the broader literature (Alfarwan, 2025).

Disclosure Statement

The authors declare that there is no conflict of interest regarding the publication of this article. No financial, personal, or professional relationships have influenced the research, analysis, or conclusions presented in this work.

Notes on Contributors

Akorfa Aku Nyatefe is a postgraduate researcher pursuing a Master's degree in Modern Educational Technology at Nanjing University of Information Science and Technology (NUIST), China. Her research interests include artificial intelligence literacy pedagogy, generative AI classroom integration, and equitable EdTech adaptation for Global South contexts. As principal author, she led the literature synthesis and framework development for this study.

nyatefeakorfa@gmail.com

Isaac Mawuena Agbogah is an academic researcher and entrepreneur with a background in Information Science and Business Administration. He holds a Bachelor's degree in Information Science from the University of Ghana and is pursuing an MBA at Nanjing University of Information Science and Technology (NUIST), China. His research interests include investment strategies, consumer behavior, database management, artificial intelligence, business and data analytics, information systems, the digital economy, e-commerce, fintech, and social media. His research explores how emerging technologies, digital platforms, and data-driven strategies influence organizational performance, consumer decision-making, and sustainable business practices.

isaacagbogah1@gmail.com

Emmanuel Akwandoh is a PhD candidate specializing in humanitarian logistics and operations research. His research focuses on applying advanced optimization and decision-support methods to complex humanitarian and disaster management problems. His work explores multi-objective optimization, evolutionary algorithms, robust optimization, and reinforcement learning for improving resource allocation, emergency response, supply chain resilience, disaster preparedness, and risk-informed decision-making. His research interests also encompass correlated disaster risks and the design of adaptive and resilient logistics systems in resource-constrained environments.

emmanuelakwandoh@outlook.com

Emmanuel Yeboah is affiliated with the Faculty of Geographical Science and Engineering, Henan University, Kaifeng, Henan Province, China. His research interests include remote sensing, geomatics, environmental monitoring, spatial analysis, and applications of geospatial technologies in resource and disaster management.

emmanuelyeboah@henu.edu.cn

Enock Peprah is a doctoral researcher in Management Science and Engineering at NUIST, working at the intersection of sustainable finance, development finance, and financial intermediation. He holds a Master of Finance from NUIST, where his thesis examined the nonlinear and spatial effects of capital market liberalization on green total factor productivity. His empirical work applies panel econometric methods, including difference-in-differences, structural equation modelling, and spatial techniques, to questions of firm financing, ESG performance, and financial development in emerging economies. A chartered member of the Institute of Chartered Accountants of Ghana, he brings eight years of public-sector financial management experience to his research on how financing instruments shape development outcomes.

202651240004@nuist.edu.cn

Sitsofe Akosua Nyatefe is an English teacher at The New Jerusalem School, Accra, Ghana. She holds a BA in Education (English) from the University of Ghana, Legon, and a DELF A2 certification in French from Alliance Française. Her teaching experience spans early childhood, basic, and secondary education, and she has served as a teaching assistant at the University of Ghana, contributing to academic instruction, student assessment, and research support. She also works as an independent proofreader and researcher, assisting with academic writing, manuscript preparation, and project development. Her research interests include educational development, language learning, adolescent cognitive and behavioral outcomes, and educational technology integration.

datruthsheltah@gmail.com

Madeakpe Gertrude Nyatefe is a fashion and design facilitator at South Labone Girls Technical Institute, Labone, and a TVET and girl child advocate. She is pursuing postgraduate studies in Education at the University of Education, Winneba. Her interests include curriculum development for TVET, sustainable fashion, and the integration of AI and digital tools into fashion design education. She seeks to bridge traditional craftsmanship and modern educational technology.

madeezklothing@gmail.com

ORCID

Akorfa Aku Nyatefe  <https://orcid.org/0009-0008-1234-3824>

Isaac Mawuena Agbogah  <https://orcid.org/0009-0000-1579-3785>

Emmanuel Akwandoh  <https://orcid.org/0009-0006-7280-4173>

Emmanuel Yeboah  <https://orcid.org/0000-0003-3838-6837>

Enock Peprah  <https://orcid.org/0009-0002-2861-1340>

Sitsofe Akosua Nyatefe  <https://orcid.org/0009-0001-6619-5037>

Madeakpe Gertrude Nyatefe  <https://orcid.org/0009-0000-8857-6124>

References

- Alfarwan, A. (2025). Generative AI use in K–12 education: A systematic review. *Frontiers in Education*, 10, Article 1647573. <https://doi.org/10.3389/feduc.2025.1647573>
- Aravantinos, S., Lavidas, K., Voulgari, I., Papadakis, S., Karalis, T., & Komis, V. (2024). Educational approaches with AI in primary school settings: A systematic review of the literature available in Scopus. *Education Sciences*, 14(7), Article 744. <https://doi.org/10.3390/educsci14070744>
- Balantekin, Y. (2026). Improving students' reading, comprehension, and reading motivation through texts generated from their prompts using ChatGPT-4: A mixed-method study. *Humanities and Social Sciences Communications*, 13, Article 9. <https://doi.org/10.1057/s41599-025-06265-x>
- Casal-Otero, L., Catalá, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K–12: A systematic literature review. *International Journal of STEM Education*, 10, Article 29. <https://doi.org/10.1186/s40594-023-00418-7>
- Druga, S., & Ko, A. J. (2021). How do children's perceptions of machine intelligence change when training and coding smart programs? In *Proceedings of the 20th ACM Interaction Design and Children Conference* (pp. 49–61). ACM. <https://doi.org/10.1145/3459990.3460712>
- Han, A., Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., & Peppler, K. A. (2024). Teachers, parents, and students' perspectives on integrating generative AI into elementary literacy

- education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Article 678). ACM. <https://doi.org/10.1145/3613904.3642438>
- Jauhiainen, J. S., & Garagorry Guerra, A. (2023). Generative AI and ChatGPT in school children's education: Evidence from a school lesson. *Sustainability*, 15(18), Article 14025. <https://doi.org/10.3390/su151814025>
- Jauhiainen, J. S., & Garagorry Guerra, A. (2024). Generative AI and education: Dynamic personalization of pupils' school learning material with ChatGPT. *Frontiers in Education*, 9, Article 1288723. <https://doi.org/10.3389/feduc.2024.1288723>
- Kızıldaş, Y. (2025). Integration of artificial intelligence (AI) into primary school students' writing skills: The impact of ChatGPT on creative writing and writing self-efficacy. *Journal of Educational Computing Research*. Advance online publication. <https://doi.org/10.1177/07356331251365187>
- Kızıldaş, Y. (2026). Artificial intelligence (AI) in primary school classrooms: ChatGPT's text summarization ability and its impact on attitudes towards summarization. *Education and Information Technologies*, 31(14), 6091–6118. <https://doi.org/10.1007/s10639-026-14029-6>
- Melsión, G. I., Torre, I., Vidal, E., & Leite, I. (2021). Using explainability to help children understand gender bias in AI. In *Proceedings of the 20th ACM Interaction Design and Children Conference* (pp. 87–99). ACM. <https://doi.org/10.1145/3459990.3460719>
- Mertala, P., & Fagerlund, J. (2024). Finnish 5th and 6th graders' misconceptions about artificial intelligence. *International Journal of Child-Computer Interaction*, 39, Article 100630. <https://doi.org/10.1016/j.ijcci.2023.100630>
- Mertala, P., Fagerlund, J., Lehtoranta, J., Mattila, E., & Korhonen, T. (2024). *Finnish primary school students' conceptions of machine learning* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2402.09582>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Ng, D. T. K., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers and Education: Artificial Intelligence*, 3, Article 100054. <https://doi.org/10.1016/j.caeai.2022.100054>

- OpenAI. (2026). *Terms of use (effective January 1, 2026)*. <https://www.openai.com/terms>
- Piaget, J., & Inhelder, B. (1969). *The psychology of the child* (H. Weaver, Trans.). Basic Books. (Original work published 1966)
- Relmasira, S. C., Lai, Y. C., & Donaldson, J. P. (2023). Fostering AI literacy in elementary science, technology, engineering, art, and mathematics (STEAM) education in the age of generative AI. *Sustainability*, 15(18), Article 13595. <https://doi.org/10.3390/su151813595>
- Rodríguez-García, J. D., Moreno-León, J., Román-González, M., & Robles, G. (2021). Evaluation of an online intervention to teach artificial intelligence with LearningML to 10–16-year-old students. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 177–183). ACM. <https://doi.org/10.1145/3408877.3432393>
- Shamir, G., & Levin, I. (2021). Neural network construction practices in elementary school. *KI – Künstliche Intelligenz*, 35(2), 181–189. <https://doi.org/10.1007/s13218-021-00729-3>
- Shamir, G., & Levin, I. (2022). Teaching machine learning in elementary school. *International Journal of Child-Computer Interaction*, 31, Article 100415. <https://doi.org/10.1016/j.ijcci.2021.100415>
- Song, Y., Kim, J., Xing, W., Liu, Z., Li, C., & Oh, H. (2025). Elementary school students' and teachers' perceptions toward creative mathematical writing with generative AI. *Journal of Research on Technology in Education*, 58(3), 615–637. <https://doi.org/10.1080/15391523.2025.2455057>
- Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K–12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795>
- Vartiainen, H., Tedre, M., & Jormanainen, I. (2023). Co-creating digital art with generative AI in K–9 education: Socio-material insights. *International Journal of Education Through Art*, 19(3), 405–423. https://doi.org/10.1386/eta_00143_1
- Vartiainen, H., Tedre, M., & Valtonen, T. (2020). Learning machine learning with very young children: Who is teaching whom? *International Journal of Child-Computer Interaction*, 25, Article 100182. <https://doi.org/10.1016/j.ijcci.2020.100182>

- Voulgari, I., Zammit, M., Stouraitis, E., Liapis, A., & Yannakakis, G. N. (2021). Learn to machine learn: Designing a game-based approach for teaching machine learning to primary and secondary education students. In *Proceedings of the 20th ACM Interaction Design and Children Conference* (pp. 593–598). ACM. <https://doi.org/10.1145/3459990.3465176>
- Yang, Y., Liu, J., Keane, T., Sun, D., & Yang, Y. (2025). Design and evaluation of ChatGPT-MWPS: An AI-enhanced learning system for improving primary students' mathematical word problem solving. *Smart Learning Environments*, 12, Article 68. <https://doi.org/10.1186/s40561-025-00419-9>
- Yim, I. H. Y. (2024). A critical review of teaching and learning artificial intelligence (AI) literacy: Developing an intelligence-based AI literacy framework for primary school education. *Computers and Education: Artificial Intelligence*, 7, Article 100319. <https://doi.org/10.1016/j.caeai.2024.100319>
- Yim, I. H. Y., & Su, J. (2025). Artificial intelligence literacy education in primary schools: A review. *International Journal of Technology and Design Education*, 35(5), 2175–2204. <https://doi.org/10.1007/s10798-025-09979-w>

Appendix

Search Log for the October 2026 Update

Searches were run on 14 October 2026 with a general web search tool. The first ten results per query were screened. Record counts beyond that were not kept.

No.	Query (as entered)	Purpose	Result
1	generative AI ChatGPT primary school elementary students empirical study 2026	Discovery	Yang et al. (2025) and Kızıldaş (2026) identified; Jauhiainen & Garagorry Guerra (2023, 2024) re-found
2	“ChatGPT” text summarization primary school fourth grade Türkiye experimental control group attitudes summarization 2026 Education Information Technologies	Discovery and details	Balantekin (2026) identified; Kızıldaş (2025) found; details for Kızıldaş (2026)
3	Title and author queries for each included study	Confirm design, sample, venue, and DOI	Sample sizes found for Han et al., Song et al., Relmasira et al.; not found for Vartiainen et al. (2023)
4	Title queries for pre-2023 studies	Confirm abstracts and DOIs	Abstracts retrieved for five of eight
5	Queries on review and framework sources (Alfarwan; Aravantinos et al.; Yim; Ng et al.; Touretzky et al.; Casal-Otero et al.)	Context and dates	Dates and scope confirmed
6	OpenAI terms of use minimum age parental consent	Age and consent conditions	OpenAI Terms of Use, effective 1 January 2026

Earlier exploratory searches for the first draft were not logged. Record counts beyond the first ten results per query were not kept.