วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น ปีที่ 49 ฉบับที่ 3 (เดือน กรกฎาคม – กันยายน 2569) หน้า 1-25
Journal of Education Khon Kaen University, Volume 49 No. 3 (July – September 2026) Page 1-25
บทความปริทัศน์ (Review Article)

วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น
https://so02.tci-thaijo.org/index.php/EDKKUJ
ดำเนินการวารสารโดย คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

Generative Artificial Intelligence in Basic Education: A Critical Integrative Review of AI Literacy, Human-AI Pedagogy, and Curriculum Redesign

Kaweechate Pia[1] & Kittirat Kasatsutorn[1]*

Faculty of Education, Thaksin University, Songkhla Campus, Thailand[1]

Received: April 19, 2026 Revised: June 10, 2026 Accepted: August 27, 2026

Abstract

Generative artificial intelligence (GenAI) has entered basic education faster than evidence, curriculum, and governance have developed. Existing discussions often combine generic artificial intelligence with GenAI, report immediate task performance as learning, and transfer findings from higher education to school-age learners. This critical integrative review examines how GenAI is being used in K-12 teaching and learning, what knowledge, skills, and attitudes it may support, which risks condition those outcomes, and what curriculum and teacher-development responses are warranted. The analytic corpus comprised 47 documents published mainly from 2020 to August 2026: 11 research syntheses, 16 empirical or contextual studies, nine policy and competency documents, and 11 foundational or methodological works. Sources were coded by educational level, AI role, learning activity, outcome domain, study design, risk, and implementation condition; empirical claims were weighted more heavily than conceptual or policy propositions. The evidence supports cautious conclusions. Structured, teacher-guided use can improve feedback, ideation, explanation, and short-term task performance, but durable learning remains uncertain and unguarded outsourcing may reduce independent performance. AI literacy therefore requires technical understanding, critical verification, ethical judgment, and learner agency rather than prompting skill alone. An author-constructed 5E application and a human-centered implementation framework position GenAI as a bounded tutor, collaborator, or assistant while preserving productive struggle and AI-free assessment checkpoints. For Thailand, adoption should be conditioned by Thai-language and cultural validation, child-data protection, rural infrastructure, and sustained teacher competence. The review concludes that curriculum redesign should optimize human-AI complementarity, not automate schooling.

Keywords: learner agency; cognitive offloading; teacher competence; child rights; glocalization

*Corresponding author. Tel.: 083 657 0419
Email address: kittirat@tsu.ac.th

Introduction

Generative artificial intelligence (GenAI) differs educationally from earlier forms of artificial intelligence because it can produce apparently original text, images, code, explanations, and feedback in response to conversational prompts. Earlier artificial intelligence in education commonly operated through relatively bounded systems, such as adaptive practice, automated scoring, prediction, or intelligent tutoring. By contrast, general-purpose GenAI can participate in many stages of schoolwork, including defining a problem, proposing an argument, drafting an answer, translating a passage, generating examples, writing code, and evaluating a product. This flexibility lowers the threshold for educational use, but it also makes the system's pedagogical role difficult to see. The same chatbot can function as a tutor that elicits reasoning, a partner that offers alternatives, or a substitute that completes the task. Educational value therefore depends less on access to the tool than on the design of the interaction.

This distinction matters because rapid adoption has encouraged an overly simple narrative: GenAI increases efficiency, therefore it improves learning. Efficiency and learning are not equivalent. A learner may complete more problems or produce a more fluent essay while the system is available, yet retain less knowledge or demonstrate weaker independent performance after assistance is removed. In a large field experiment with nearly 1,000 secondary mathematics students, unrestricted access to a general-purpose interface improved practice performance but was followed by lower unaided test performance; a tutor interface with pedagogical guardrails largely mitigated that penalty (Bastani et al., 2025). The finding does not establish that all GenAI use harms learning. It shows that interface design, task structure, and the timing of assistance can reverse an apparently positive result. The Organisation for Economic Co-operation and Development (OECD, 2026) similarly distinguishes task performance from learning and argues that pedagogically guided use is a necessary condition for educational benefit.

The evidence base in basic education is growing, but it remains fragmented. A recent K-12 systematic review identified 45 studies published from 2020 to 2024 and reported gains in achievement, motivation, and personalization alongside inaccurate content, dependence, and privacy risks (Zhang, Lai, et al., 2026). Another review of 30 K-12 studies found that only nine used quasi-experimental designs, psychological outcomes were measured more often than objective learning, and the literature underrepresented rural learners, students with disabilities, and African contexts (Alfarwan, 2025). Zhang and Tur's (2024) earlier review contained only 13 K-12 publications, illustrating how quickly the field has expanded. A broader mapping review likewise describes a rapidly expanding, heterogeneous literature that mixes educational settings and outcomes (Yusuf et al., 2024). At the review-of-reviews level, Zhang, Deng, et al. (2026) synthesized 35 reviews and found recurring methodological inconsistency, weak theoretical grounding, limited outcome validation, and few evaluated mitigation strategies. These findings counsel against broad claims that GenAI has already transformed school learning.

Three additional gaps require attention. First, conceptual papers, policy guidance, perception surveys, short classroom demonstrations, and causal studies are often discussed as if they provide equivalent evidence. They do not. A teacher's belief that a tool may personalize instruction is important for

implementation but does not establish a learning effect. Second, studies frequently conflate knowing about artificial intelligence, using a chatbot, and learning disciplinary content with a chatbot. These are different curricular aims. Third, global recommendations are often transferred to national settings without testing language performance, cultural alignment, infrastructure, or child-protection requirements. In Thailand, teachers recognize opportunities for personalized learning and efficiency but also identify concerns about accuracy, ethics, readiness, and access (Tongchai & Malakul, 2026). Thai-language benchmarks further show that general reasoning performance and culturally appropriate performance are distinct properties: model size or multilingual capability does not guarantee sensitivity to Thai norms and linguistic context (Kim et al., 2025).

The central problem is therefore not whether schools should be “for” or “against” GenAI. The more useful question is how basic education can preserve human agency, disciplinary thinking, and equitable participation while using selected AI capabilities. This review addresses that problem through four questions:

1. How is GenAI being used as a pedagogical tool in basic education, and under what conditions does it support learning rather than task substitution?
2. What knowledge, skills, and attitudes (KSA) are associated with responsible GenAI use, and how strong is the supporting evidence?
3. What ethical, cultural, equity, privacy, and cognitive risks are especially consequential for school-age learners?
4. What implications follow for curriculum design, classroom assessment, teacher development, and education policy, particularly in Thailand?

The article makes three contributions. It separates evidence of immediate performance from evidence of durable learning; it integrates AI literacy, hybrid human-AI learning, cognitive offloading, and child-rights perspectives; and it translates the synthesis into an author-constructed 5E application and a human-centered curriculum framework. The intended result is not a universal implementation recipe. It is a set of design propositions whose strength and limits can be examined by educators, researchers, and policymakers.

■ Conceptual Foundations

AI literacy as critical, practical, and ethical competence

AI literacy is broader than the ability to write effective prompts. Long and Magerko (2020) describe competencies that allow people to recognize, understand, evaluate, and interact with AI. Ng et al. (2021) organize the construct around knowing and understanding, using and applying, evaluating and creating, and ethical issues. UNESCO's student framework extends this orientation through a human-centered mindset, ethics of AI, AI techniques and applications, and AI system design, with progression from understanding to application and creation (Miao & Shiohira, 2024). Its teacher framework adds AI pedagogy and professional learning and stresses that competence develops across acquisition, deepening, and creation stages (Miao & Cukurova, 2024).

These frameworks share two principles. First, AI literacy includes judgment about when not to use AI. A student who can generate a fluent answer but cannot verify evidence, identify uncertainty, or explain the reasoning is not fully AI literate. Second, ethics is not a final lesson added after technical skills. It should shape data selection, prompt design, evaluation of outputs, attribution, disclosure, and decisions about human responsibility throughout the learning process. Ma et al.'s (2025) synthesis of 68 publications on K-12 AI ethics education reaches a similar conclusion: responsible AI literacy should permeate cognitive, affective, and behavioral learning rather than remain a detached topic.

For this review, AI literacy is operationalized as four connected capacities: (a) conceptual knowledge of how GenAI is trained, generates outputs, and fails; (b) practical skill in framing tasks, interacting iteratively, and integrating outputs; (c) evaluative skill in checking accuracy, provenance, bias, and fitness for purpose; and (d) ethical agency in deciding, disclosing, and accepting responsibility for use. This definition distinguishes literacy from mere adoption and provides a basis for interpreting KSA outcomes.

Human-AI complementarity and learner agency

Human-centered education treats AI as a means rather than an educational end. Miao and Holmes (2023), the U.S. Department of Education (2023), and OECD (2026) all emphasize that systems should augment human judgment and relationships. Molenaar's (2022) account of hybrid human-AI learning similarly positions human and machine capabilities as complementary, while Kong and Yang (2024) connect GenAI use to domain learning and self-regulation. The relevant design question is which agent should perform which cognitive function, at what time, with what visibility and accountability.

Three bounded roles are useful. As a tutor, GenAI asks questions, supplies hints, diagnoses an explanation, or adapts practice without revealing a complete solution prematurely. As a collaborator, it proposes alternatives that learners compare, contest, and improve. As an assistant, it handles lower-level operations, such as reformatting student-authored notes or producing accessibility supports, after the learner has made the substantive decisions. The same roles become educationally hazardous when boundaries disappear: tutoring becomes answer delivery, collaboration becomes ghostwriting, and assistance becomes unobserved substitution. Learner agency is preserved when students set goals, make consequential choices, explain decisions, and remain able to perform essential tasks without the system.

Dewey's (1916) account of learning through inquiry remains relevant because experience is educative only when it supports reflection, continuity, and further intelligent action. GenAI can broaden the materials available for inquiry, but it cannot replace the learner's encounter with uncertainty, evidence, and consequences.

From pre-GenAI AIED to generative AI: continuity and departure

Holmes et al. (2019) provide a useful historical baseline for distinguishing contemporary GenAI from earlier artificial intelligence in education (AIED). Their book considers both what students should learn in an AI-shaped society and how AI might alter teaching and learning. Its tentative functional taxonomy groups AIED applications as primarily *student teaching*, *student supporting*, or *teacher supporting*, while cautioning that the categories overlap and may converge in multi-capable systems (Holmes et al., 2019, pp. 164–168).

The book also questions the robustness and scalability of efficacy claims and identifies unresolved concerns about accuracy, bias, learner choice, privacy, teachers' roles, and ethics (Holmes et al., 2019, pp. 169–180). These cautions remain relevant; however, the technologies examined were mainly bounded systems such as intelligent tutors, dialogue-based tutors, exploratory learning environments, automated writing evaluation, and learning analytics—not today's general-purpose generative models.

The present review is therefore neither a summary of nor a substitute for the 2019 book. It treats the book as a pre-GenAI conceptual comparator and examines which propositions remain continuous, which require revision, and which new boundary conditions arise when one conversational system can generate content and shift among instructional roles. Table 1 makes this relationship explicit. The extension rests on a 47-document corpus focused mainly on 2020–August 2026 and on basic education; it adds design-sensitive evidence about assisted versus independent performance, an operational AI-literacy framework, child-rights safeguards, and conditions specific to Thailand.

Table 1. *Continuity with and departure from Holmes et al. (2019)*

Dimension	Holmes et al. (2019): pre-GenAI baseline	Extension in the present review
Focal technologies	Predominantly bounded AIED applications, including intelligent and dialogue-based tutoring, exploratory environments, automated writing evaluation, learning analytics, and teacher-support tools	General-purpose GenAI capable of producing text, images, code, explanations, and feedback across many stages of one learning task
Educational roles	Tentative, overlapping functions of student teaching, student supporting, and teacher supporting	Bounded tutor, collaborator, and assistant roles whose educational value depends on when assistance enters and which learner processes remain visible
Evidence stance	Cautious account of promise; limited robust evidence of efficacy at scale and concern that AIED may be oversold	Critical integrative synthesis of 47 documents, with claims weighted by design and with immediate task performance separated from durable, independent learning
Learning and agency	Personalization and augmentation are possible, but choice, assumptions embedded in systems, and consequences for learners require scrutiny	Productive struggle, cognitive offloading, fading support, learner decision-making, and AI-free transfer checkpoints are explicit design requirements
Curriculum assessment	Addresses what learners should know in an AI-shaped world and how AIED may modify teaching, learning, and	Defines conceptual, practical, evaluative, and ethical AI literacy and applies a human-first, bounded-AI, human-final sequence through an

	continuous assessment	author-constructed 5E design
Ethics and context	Raises accuracy, bias, privacy, teacher-role, and emerging AIED ethics questions	Adds child-data minimization, meaningful non-AI alternatives, procurement accountability, Thai-language and cultural validation, rural access, and revalidation after model updates

Note. The middle column synthesizes Holmes et al. (2019, pp. 164–180). “Extension” identifies the analytical contribution of this review; it does not imply that the proposed curriculum designs have already been validated as interventions.

Cognitive offloading, productive struggle, and assessment validity

Cognitive offloading occurs when people use external actions or tools to reduce internal cognitive demand (Risko & Gilbert, 2016). Offloading is not inherently harmful. Notes, calculators, diagrams, and search engines can release capacity for higher-order work. The educational issue is whether the offloaded process is itself a learning objective or a prerequisite for later reasoning. If a novice delegates explanation, retrieval, or problem representation before building a mental model, short-term fluency may conceal weak learning. If an experienced learner delegates routine formatting after constructing an argument, the same tool may improve efficiency without undermining the target competence.

This distinction leads to two design principles. Productive struggle should precede high-level AI assistance when the struggle is necessary to build knowledge or strategy. Independent checkpoints should follow assisted practice when the curriculum claims that learners have acquired transferable competence. Bastani et al. (2025) provide direct evidence for this distinction in secondary mathematics; broader meta-analytic evidence suggests positive average effects of ChatGPT-supported instruction, but variations in educational level, duration, comparison condition, and outcome measurement limit direct inference to all K-12 contexts (Deng et al., 2025). Assessment must therefore record the process and test unaided performance, not only judge a polished AI-assisted product.

■ Review Design

Rationale and scope

This article uses a critical integrative review, a design suited to an emerging field that contains empirical studies, research syntheses, conceptual frameworks, and policy documents (Torraco, 2016; Whittemore & Knafl, 2005). It does not claim to be a new exhaustive systematic review or meta-analysis. That distinction is deliberate. Several recent systematic reviews already provide database-level coverage, but their scopes and study sets overlap. The present purpose is to compare what those reviews establish, connect their findings to influential empirical studies and authoritative frameworks, expose contradictions, and derive context-sensitive curriculum propositions.

The unit of analysis was the published document, not a human participant. No learners or teachers were recruited and no personal data were collected; consequently, “participants” and human-subject

procedures are not applicable to this review. When sample sizes are reported, they refer to participants in the primary studies being discussed, not to participants in the present article.

Source identification and eligibility

The corpus was developed through a structured, purposive process completed in August 2026. Recent K-12 systematic reviews and meta-reviews served as anchor documents. Backward and forward citation chaining was then used to locate influential empirical studies, theoretical accounts, and implementation evidence. Targeted searches were used to add Thai studies and official guidance from UNESCO, UNICEF, OECD, the U.S. Department of Education, and the International Telecommunication Union. Foundational works published before 2020 were retained only when they defined a core construct or model used in the synthesis.

Documents were eligible when they met at least one of four functions: (a) a systematic, mapping, or meta-analytic synthesis concerning GenAI, AI literacy, or AI ethics in education; (b) an empirical or implementation study involving primary or secondary education, teachers, or school-age learners; (c) an authoritative policy, competency, child-rights, or national-context document; or (d) a foundational theoretical or methodological source needed to interpret the evidence. Higher-education-only studies were excluded from claims about basic education unless they appeared in a broader synthesis and were explicitly used only for conceptual comparison. Vendor marketing, unreviewed opinion pieces, and sources without a transparent evidentiary or institutional basis were excluded. The main publication window was January 2020 to August 2026, with six pre-2020 foundational or methodological works retained.

The final analytic corpus comprised 47 documents. Table 2 shows the composition and the function of each source type. These categories are not equivalent evidence tiers: policy documents establish norms and priorities, while causal learning claims require empirical designs with appropriate comparison and outcome measures.

Table 2. *Composition and use of the analytic corpus*

Source group (n)	Primary role in the synthesis	Evidentiary treatment
Systematic, mapping, exploratory, and meta-analytic reviews (n = 11)	Map study characteristics, recurring outcomes, gaps, and cross-study patterns	Highest weight for field-level patterns; overlap between reviews prevents simple aggregation of counts
K-12 empirical and contextual studies (n = 16)	Test or document learning activities, perceptions, implementation, language/cultural performance, and risks	Weight varied by design, sample, duration, outcome validity, and independence of assessment
Policy, competency, child-rights, and infrastructure	Define public-interest requirements, competencies,	Used as normative or implementation guidance, not as

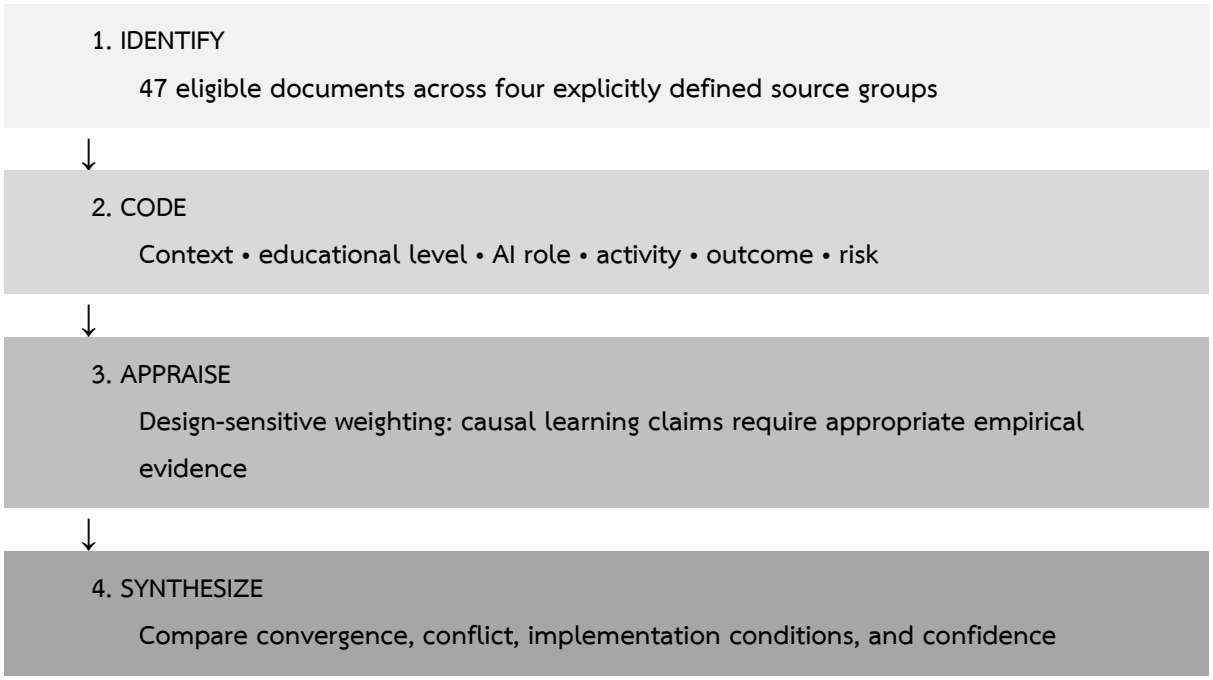
documents (n = 9)	governance duties, and contextual constraints	causal evidence of learning effects
Foundational conceptual and methodological works (n = 11)	Define AI literacy, human-AI complementarity, cognitive offloading, inquiry, 5E, and integrative review method	Used to interpret evidence and construct propositions; not counted as intervention evidence
Total corpus (n = 47)		

Coding, appraisal, and synthesis

Coding was conducted in two passes. The descriptive pass extracted year, educational level, country or region, AI type, pedagogical role, learning activity, outcome domain, study design, sample or corpus size when available, and reported limitations. The interpretive pass coded four cross-cutting dimensions aligned with the review questions: pedagogical use; KSA outcome; ethical, cognitive, cultural, equity, and privacy risk; and curriculum or teacher-development implication. Evidence, interpretation, and recommendation were kept in separate fields of the analytic matrix to reduce the risk of converting an author's proposal into an observed result.

Source quality was appraised through design-sensitive questions rather than a single score across incomparable document types. Reviews were examined for a defined search process, eligibility criteria, study appraisal, and transparent synthesis. Empirical studies were examined for an appropriate comparison, sample and context description, intervention duration, validated or independent outcomes, and limitations. Policy documents were examined for institutional authority, transparency, rights orientation, and relevance to basic education. Conceptual works were examined for definitional clarity and explanatory value. Findings were classified as relatively stronger when supported by convergent reviews or independent performance measures; as moderate when supported by several short-term or context-bound studies; and as emerging when based mainly on perception data, conceptual proposals, or single cases.

No inter-rater reliability coefficient is reported because the review was not designed as a dual-reviewer systematic review. This is a methodological limitation, and the article therefore avoids claims of exhaustive coverage or statistically pooled effects. Its rigor rests instead on explicit scope, source-role separation, design-sensitive weighting, contradiction analysis, and traceable citations.



Evidence, interpretation, and recommendation were retained as separate analytic fields

Figure 1. *Author-constructed workflow for the critical integrative review. The workflow is not a PRISMA flow diagram and does not imply an exhaustive systematic search.*

Results

Profile and limits of the evidence base

The review literature shows rapid growth after the public release of conversational GenAI, but the apparent volume overstates the maturity of school-level evidence. The 30 studies reviewed by Alfarwan (2025), the 45 reviewed by Zhang, Lai, et al. (2026), and the 13 reviewed by Zhang and Tur (2024) partly cover the same early studies. These counts must not be summed. The meta-review of 35 reviews by Zhang, Deng, et al. (2026) similarly represents a synthesis layer, not 35 independent bodies of primary evidence. Across these reviews, recurring limitations include short interventions, convenience samples, self-reported engagement or intention, weak comparison conditions, and insufficient measurement of durable or transferable learning.

The distribution of studies is also uneven. Alfarwan (2025) found ChatGPT in 21 of 30 K-12 studies and only nine quasi-experimental designs. Many outcomes concerned motivation, perceptions, or technology acceptance rather than independently assessed disciplinary learning. Studies were concentrated in parts of Asia and Europe; rural settings, Africa, special education, and sustained school-wide implementation were limited. Zhang, Lai, et al. (2026) likewise concluded that benefits coexist with inaccurate content, dependence, and privacy concerns and that long-term evidence is scarce. Consequently, confidence is greatest for bounded, near-term functions such as generating examples or receiving feedback and weakest for claims about long-term achievement, developmental effects, equity, and system-level transformation.

The emerging Thai evidence illustrates both value and uncertainty. In a survey of 416 Thai teachers, respondents identified instructional and efficiency opportunities while reporting concerns about accuracy, ethics, readiness, and access (Tongchai & Malakul, 2026). A study of 121 Thai secondary students reported generally positive perceptions of GenAI-supported summative assessment but also concerns about overreliance, accuracy, and the need for explicit ethical guidance (Nantaburom, 2026). These studies are useful for understanding acceptability and implementation conditions; neither design by itself demonstrates a causal improvement in learning. The distinction between perception and effect is therefore maintained throughout the synthesis.

Pedagogical uses: tutoring, collaboration, and assistance

A K-12 review organized through presage, process, and product relationships likewise indicates that instructional goals, learning activities, AI roles, and outcomes must be interpreted together rather than treated as isolated tool effects (Lin & Tan, 2025).

GenAI as a bounded tutor

The most educationally defensible tutor role is contingent support: the system offers a hint, asks a diagnostic question, or provides feedback responsive to a learner's attempt. This role can increase opportunities for practice when teacher time is limited. In physics and science contexts, small studies have used ChatGPT to generate explanations, questions, or feedback, with learners often reporting usefulness and engagement (Alneyadi & Wardat, 2023; Bitzenbauer, 2023). A classroom lesson study similarly illustrates the value of guided inquiry into a chatbot's responses rather than treating the system as an autonomous authority (Jauhiainen & Guerra, 2023). In programming, structured GenAI support has been associated with gains in problem-solving or learning experience, although results remain sensitive to prior knowledge and task design (Yang et al., 2025). A human-centered K-12 framework proposed by Kong and Yang (2024) similarly organizes GenAI support around domain learning and self-regulation rather than unrestricted answer production.

However, a conversational interface is not automatically a tutor. A tutor must preserve the learner's cognitive work. Bastani et al. (2025) provide the clearest warning: unrestricted assistance produced large practice gains but weaker unaided performance, whereas a guardrailed interface designed to support rather than answer substantially reduced the harm. The result is consistent with cognitive offloading theory (Risko & Gilbert, 2016). When the system performs problem representation, strategy selection, and explanation before the learner attempts them, the learner may experience fluency without constructing knowledge. The tutor role is therefore supported with moderate confidence only when assistance is delayed, graduated, and followed by independent performance.

GenAI as a dialogic collaborator

As a collaborator, GenAI can widen a learner's option space. It can propose alternative explanations, simulate a counterargument, vary a problem, or generate contrasting examples for evaluation. In elementary or secondary literacy, teachers, parents, and students have identified opportunities for idea generation and feedback alongside concerns about authorship, accuracy, and age-appropriate use (Han et

al., 2024; Hays et al., 2024). In creative and arts-based activity, learners can compare human and machine choices and examine how prompts, training data, and platform constraints shape an artifact (Vartiainen et al., 2023). Dialogic designs in secondary education have also used GenAI outputs as objects for critique rather than final answers (Tang et al., 2024).

Collaboration is educationally meaningful when the learner contributes an independent position and must justify selection, rejection, or revision. A useful pattern is “human first, AI second, human final”: the student records an initial idea; consults the system for alternatives or critique; then produces and defends a revised response with a disclosure of AI use. The pattern makes change visible and gives teachers evidence of reasoning. By contrast, prompting a system to produce an initial essay or complete design and then making surface edits provides weak evidence of the learner's competence. The literature supports ideation and dialogic critique as promising uses, but evidence for durable creativity or critical thinking remains emerging because studies often measure perceptions or product quality without an independent transfer task.

GenAI as an assistant for access and teacher work

GenAI can assist with translation, reading-level adjustment, example generation, lesson planning, and the production of alternative representations. Kindergarten and school teachers perceive possibilities for differentiated materials and administrative efficiency but also report uncertainty about reliability, appropriateness, and pedagogical knowledge (Su & Yang, 2024; Tongchai & Malakul, 2026). Science and mathematics teachers similarly request professional support that combines tool operation with subject-specific and ethical judgment (Cheah & Kim, 2025).

Assistance may improve access, especially when students require language scaffolds or multiple representations. Yet automated simplification can remove disciplinary precision, translation can distort local meaning, and personalization based on student data can create privacy or profiling risks. The assistant role therefore requires teacher review and a clear data-minimization rule. It should not involve uploading identifiable student work, disability information, behavioral records, or assessment data to an unapproved service. The evidence supports time savings and expanded material generation more strongly than it supports learning improvement; claims should be framed accordingly.

Knowledge, skills, and attitudes

The KSA framework is used here as an analytic organizer, not as a validated causal model. Its value is to prevent curriculum from reducing AI literacy to one type of outcome. Table 3 summarizes the evidence, conditions, and confidence for each domain.

Table 3. *Evidence synthesis for learner outcomes*

Domain and confidence	Potentially supported outcomes	Required conditions and key limitations
Knowledge — Moderate (near-term); low (durable retention)	AI concepts and limitations; disciplinary explanations; recognition of uncertainty and provenance	Explicit instruction, source comparison, retrieval and explanation before assistance, and independent checks are required; fluent errors and shallow encoding remain important limitations
Skills — Moderate for bounded practices; emerging for transfer	Prompt framing; questioning; comparison; revision; data and source verification; human-AI collaboration	Human-first attempts, iterative feedback, process logs, transfer tasks, and teacher modeling are required; task outsourcing, inequitable access, and product-only assessment can invalidate conclusions
Attitudes and agency — Emerging	Curiosity; confidence; ethical responsibility; willingness to question AI; calibrated trust	Age-appropriate ethics, visible uncertainty, choice, reflection, and discussion of affected communities are required; perception evidence currently exceeds behavioral evidence
Responsible AI literacy — Moderate (framework consensus); emerging (measured behavior)	Integration of conceptual, practical, evaluative, and ethical competence	Ethics should permeate activities, bias and privacy should be audited, and human accountability should remain visible; isolated compliance lessons are insufficient

Knowledge

GenAI can provide multiple explanations and examples at low marginal cost, which may help learners connect representations or receive immediate clarification. Reviews report positive achievement outcomes, but most interventions are short and often compare GenAI-supported activity with business-as-usual instruction without isolating the active mechanism (Alfarwan, 2025; Zhang, Lai, et al., 2026). Broader experimental meta-analysis reports an average positive effect of ChatGPT-supported learning, yet heterogeneity and inclusion of higher education limit direct generalization to basic education (Deng et al., 2025).

Knowledge gains are most plausible when students retrieve or construct an explanation before seeing generated content, compare the output with authoritative sources, and revisit the concept without assistance. The opposite sequence can induce an illusion of knowing. A polished explanation may feel familiar and understandable while leaving the learner unable to reconstruct it. Accordingly, GenAI should often follow, not precede, initial retrieval, prediction, or problem representation.

Knowledge about AI itself is also essential. Students should understand that a large language model predicts patterns in data rather than consulting a stable internal truth database; that outputs can vary with

prompts and system versions; that confidence in wording is not calibrated confidence in accuracy; and that training data can encode historical and representational bias. These concepts can be taught at age-appropriate levels without requiring advanced programming. AI literacy reviews and UNESCO frameworks converge on the need to combine conceptual understanding with evaluation and ethics (Casal-Otero et al., 2023; Miao & Shiohira, 2024; Ng et al., 2021).

Skills

The most frequently discussed skills are prompting, questioning, revision, evaluation, and self-regulation. Prompting has instructional value when treated as precise communication and iterative inquiry, not as a collection of tricks. Learners can define a goal, state constraints, ask for alternatives, request evidence, test consistency, and refine the interaction. Yet prompting alone can become a procedural shortcut that masks weak domain knowledge. A learner cannot reliably evaluate an answer in a domain they do not understand.

Higher-value skills therefore involve verification and integration. Students should trace claims to accessible sources, compare output with data or texts, detect missing assumptions, explain revisions, and decide whether AI is appropriate for the task. In scientific inquiry, learners may use GenAI to generate hypotheses but should still design observations, inspect evidence, and justify conclusions (Min et al., 2025). In creative mathematical writing, GenAI may stimulate variation, but assessment should examine mathematical validity and the learner's transformation of suggestions (Song et al., 2025). In programming, code generated by AI should be tested, explained, and modified rather than merely executed (Yang et al., 2025).

Self-regulation is a promising but conditional outcome. GenAI can help learners set goals, obtain feedback, or reflect, and classification tools have been proposed to align prompts with self-regulated learning and self-determination (Chiu, 2024). The same immediacy can weaken monitoring if learners accept the first plausible answer. Effective designs externalize self-regulation through planning notes, prompt-and-response logs, confidence ratings, error analysis, and a final statement of what the learner—not the system—decided.

Attitudes and ethical agency

Many school studies report enthusiasm, usefulness, confidence, or intention to use. These attitudes affect adoption but should not be interpreted as evidence of competence. Positive perceptions may coexist with overtrust or inaccurate beliefs. Conversely, concern about AI does not necessarily indicate rejection; it can reflect appropriate caution. The educational target is calibrated trust: learners should know which outputs can be used provisionally, which require verification, and which decisions must remain human.

Ethical agency includes recognizing affected parties, anticipating consequences, and accepting responsibility. Ma et al. (2025) found that K-12 AI ethics education addresses cognitive, affective, and behavioral domains but remains uneven across regions and responsible-AI principles. Learners need opportunities to examine representational harm, data consent, intellectual contribution, environmental and

labor implications, and the unequal consequences of error. A classroom rule such as “cite AI” is insufficient if students do not understand why provenance, accountability, and affected communities matter.

Risks and boundary conditions

Hallucination, bias, and the appearance of authority

GenAI can produce fabricated facts, sources, and quotations in fluent language. This is not a minor technical defect in school settings, where students may lack the background knowledge needed to detect error. Verification must therefore be designed into the task. Students should identify checkable claims, consult authoritative sources, record corrections, and explain why a source is credible. Teachers should not rely on generated answer keys or feedback without review.

Bias is both statistical and social. Educational algorithms can reproduce disparities through data, labels, proxies, design choices, and unequal error rates (Baker & Hawn, 2022). GenAI adds representational risk because it produces culturally situated language and examples. A bias audit should ask whose language and experience are treated as default, which groups are absent or stereotyped, and who bears the cost of an incorrect recommendation. The audit must examine output in the language and context of actual use, not assume that English-language performance transfers.

Cognitive offloading and dependency

Dependency is not measured by frequency of use alone. It occurs when access to the tool reduces the learner's capacity or willingness to perform essential processes independently. The Bastani et al. (2025) experiment demonstrates that impressive assisted performance can conceal a learning penalty. Reviews also repeatedly identify overreliance as a concern (Alfarwan, 2025; Zhang, Lai, et al., 2026). The appropriate response is not a universal ban but staged access: unaided attempt, bounded assistance, explanation and source checking, followed by delayed unaided performance. Schools should monitor the gap between assisted and independent outcomes.

Privacy, child rights, and surveillance

School-age learners cannot be treated as ordinary adult consumers. UNICEF (2025) identifies requirements that include safety, privacy, nondiscrimination, transparency, accountability, inclusion, and support for children's development and agency. Miao and Holmes (2023) similarly emphasize data protection, age appropriateness, human oversight, and institutional validation. These principles imply practical controls: use approved systems; minimize data; prohibit entry of personally identifiable or sensitive student information; provide meaningful notice and alternatives; document retention and model-training settings; and maintain a non-AI route for participation.

The use of GenAI for behavioral prediction, emotion inference, high-stakes placement, discipline, or automated grading deserves especially strict scrutiny. Such applications can create opaque records and shift authority away from educators and families. Human review is necessary but not sufficient if the underlying use is disproportionate or invalid. Governance should begin by asking whether the system should be used at all.

Equity, infrastructure, and access

GenAI may expand access to feedback, but it may also deepen differences in devices, connectivity, paid features, language support, and adult guidance. The Thai digital divide includes geographic and socioeconomic dimensions in connectivity and digital participation (International Telecommunication Union, 2022), while global evidence shows that educational technology can reproduce exclusion when access, governance, and context are ignored (UNESCO Global Education Monitoring Report Team, 2023). A curriculum that assumes continuous access to a high-performing commercial service can disadvantage rural schools and families and can become obsolete when pricing or product policies change.

Equity therefore requires more than distributing accounts. Schools need shared-device and low-bandwidth designs, non-AI alternatives with equivalent learning value, accessible interfaces, local-language evaluation, and teacher capacity to support learners who receive less help outside school. Procurement should compare the total conditions for effective use, including connectivity, training, privacy review, support, and exit options, not only model capability.

Thai language and cultural alignment

Glocalization means adapting global principles through local language, culture, curriculum, and institutional conditions. It is not the superficial insertion of Thai examples into an imported tool. Kim et al. (2025) developed Thai-H6 and ThaiCLI because mainstream benchmarks underrepresent Thai reasoning and cultural-linguistic performance. Their findings indicate that capability on general Thai tasks and culturally aligned judgment do not move in identical ways. This has direct educational implications: a model may produce grammatical Thai yet mishandle politeness, social relationships, local history, regional identities, or culturally sensitive dilemmas.

Schools should therefore validate tasks with Thai-language experts and relevant communities before routine deployment. Evaluation should include standard Thai, regional and minority language needs where appropriate, age-level readability, culturally sensitive scenarios, and known curricular content. Teachers need authority to reject or revise generated materials. Local validation should be repeated after model updates because system behavior is not stable over time.

An author-constructed 5E application

The 5E instructional model—Engage, Explore, Explain, Elaborate, and Evaluate—provides a useful sequence for preserving inquiry when GenAI is available (Bybee et al., 2006). Table 4 is an author-constructed application derived from the evidence in this review; it is not presented as a protocol found intact in the included studies. Elaborate and Evaluate are kept separate because transfer and judgment serve different functions.

Table 4. *Author-constructed 5E application for bounded GenAI use*

5E phase and human learning purpose	Bounded GenAI role	Required evidence and safeguard
Engage — Elicit prior knowledge, curiosity, and an authentic local problem	Offer contrasting claims or a provocative scenario after learners make an initial prediction	Record pre-AI prediction; teacher checks age and cultural appropriateness
Explore — Investigate patterns, evidence, and possible explanations	Generate testable alternatives, simulations, or questions—not a final explanation	Learners collect or inspect evidence; prompts and uncertain claims are logged
Explain — Construct and communicate a disciplinary explanation	Critique a student explanation, identify gaps, or provide a contrasting account	Student explains first; claims are checked against authoritative sources; revisions are justified
Elaborate — Transfer understanding to a new context or locally meaningful problem	Propose constraints, stakeholders, or alternative cases for application	Learners adapt—not copy—the response; performance is tested in a new context
Evaluate — Determine independent learning, process quality, and ethical judgment	Support self-questioning or formative feedback, but not produce the assessed response	AI-free checkpoint; process log; source/bias audit; disclosure and reflection on human decisions

The sequence implements "human first, AI second, human final." It prevents Engage from becoming passive consumption, Explore from becoming answer retrieval, and Explain from becoming a paraphrase of generated text. Elaborate requires genuine transfer to a novel or local context. Evaluate uses independent evidence so that teachers can distinguish an improved product from acquired competence. The design can be adapted to subject and developmental level, but its safeguards should not be removed merely for convenience.

■ Discussion

What the evidence supports—and what it does not

Across reviews, policy guidance, and empirical studies, there is broad agreement that GenAI can expand access to examples, feedback, dialogue, and material generation. There is also broad agreement that outputs can be inaccurate or biased and that learners need verification, ethical guidance, and teacher mediation (Miao & Holmes, 2023; OECD, 2026; Wang et al., 2024; Zhang, Lai, et al., 2026). The principal disagreement concerns the strength of learning claims. Reviews often describe positive outcomes, whereas design-sensitive analysis reveals that many outcomes are immediate, self-reported, or measured while

assistance remains available. Bastani et al. (2025) show why this matters: better assisted work and better learning can diverge.

Variation is likely produced by at least five moderators. First, educational level affects prior knowledge and the ability to evaluate output; younger learners require more constrained interfaces and adult mediation. Second, subject matter changes the consequences of error and the value of externalization. A generated story idea and a generated chemical safety instruction cannot be governed identically. Third, AI role matters: a hint-based tutor preserves more learner work than a completion engine. Fourth, study design matters: product quality, enjoyment, and independent transfer are different outcomes. Fifth, context matters: language performance, teacher competence, device access, and cultural fit condition what an intervention actually delivers.

The synthesis therefore rejects both technological determinism and blanket prohibition. GenAI is not an autonomous cause of educational transformation. It is a configurable participant in an activity system whose effects depend on curriculum, interface, teacher judgment, assessment, governance, and learner history. The relevant unit of design is not the tool alone but the sequence of human and machine actions.

From tool adoption to curriculum redesign

Curriculum redesign should begin with valued human capability and work backward. For each task, designers should identify the disciplinary thinking students must learn, the processes that can be supported without being displaced, and the evidence needed to demonstrate independent competence. GenAI use is justified when it increases feedback, variation, access, or reflective comparison while preserving those processes. It is not justified merely because it makes a product faster.

This orientation changes the KSA model. Knowledge must include disciplinary understanding and an accurate mental model of AI limitations. Skills must include inquiry, verification, collaboration, and transfer—not prompting alone. Attitudes must include calibrated trust, intellectual humility, responsibility, and willingness to preserve human agency. These domains interact. Verification skill depends on knowledge; ethical action depends on recognizing uncertainty and affected parties; agency depends on opportunities to make and defend decisions.

Assessment is the leverage point. If schools grade only polished products, students and teachers are incentivized to optimize output regardless of learning. A stronger system triangulates three forms of evidence: an AI-assisted product, a process record showing prompts, sources, corrections, and decisions, and an independent performance or oral defense. Disclosure should be proportionate and educative rather than punitive. Students need clear examples of permitted assistance, required acknowledgment, and prohibited substitution. Teachers also need procedures for uncertainty; AI-detection scores should not be treated as proof of misconduct because such tools can be unreliable and can produce inequitable consequences.

Teacher competence and professional agency

Teacher development cannot be reduced to a one-off workshop on prompting. UNESCO's teacher framework includes human-centered mindset, ethics, foundations and applications, AI pedagogy, and

professional learning (Miao & Cukurova, 2024). The Thai teacher evidence similarly indicates that perceived opportunity coexists with readiness, access, accuracy, and ethical concerns (Tongchai & Malakul, 2026). Effective development should therefore be subject-specific, iterative, and connected to assessment and governance.

Teachers need at least five practical capabilities: to test a model's output against disciplinary standards; to design bounded roles and fading assistance; to recognize privacy and bias risks; to collect evidence of independent learning; and to adapt activities to language, culture, age, and access. Professional learning communities can conduct lesson studies in which teachers compare an AI-assisted task with a non-AI alternative, inspect student reasoning, and revise safeguards. Teachers should also participate in procurement and policy decisions because they see where system assumptions conflict with classroom reality.

Professional agency includes the right not to use a tool when evidence, infrastructure, or data protection is inadequate. Human oversight should not become nominal approval of machine-produced decisions. Teachers must have time, authority, and institutional support to exercise judgment.

A glocal agenda for Thailand

Thailand should adopt internationally grounded child-rights and competency principles while testing their local enactment. Four conditions are particularly important. First, Thai-language and cultural validation should precede curricular dependence. Benchmarks such as Thai-H6 and ThaiCLI show why grammatical output is insufficient (Kim et al., 2025). Evaluation should include regional, minority, and culturally sensitive cases and should involve educators and communities affected by errors.

Second, infrastructure policy should treat connectivity, devices, technical support, and non-AI alternatives as part of pedagogy. The digital divide cannot be solved by issuing a national recommendation that presumes uniform access, and effective digital education requires an ecosystem rather than a stand-alone tool (International Telecommunication Union, 2022; OECD, 2023). Pilot designs should include rural and under-resourced schools from the beginning, not after an urban model has been declared successful.

Third, data governance should be child-centered. Approved tools, minimum-data rules, consent and notice, retention limits, incident procedures, and access to a meaningful alternative should be specified before classroom use. High-stakes decisions and sensitive records require stricter limits than low-risk brainstorming. Procurement contracts should address whether prompts and outputs are retained or used for model training.

Fourth, evidence should be generated locally. Thai studies should measure independent learning after AI access is removed, not only attitudes or assisted product quality. Longer follow-up, subject-specific outcomes, teacher workload, equity effects, and implementation costs are needed. Comparative studies should document model version, language, prompt conditions, teacher mediation, and the non-AI comparison so that results remain interpretable as systems change.

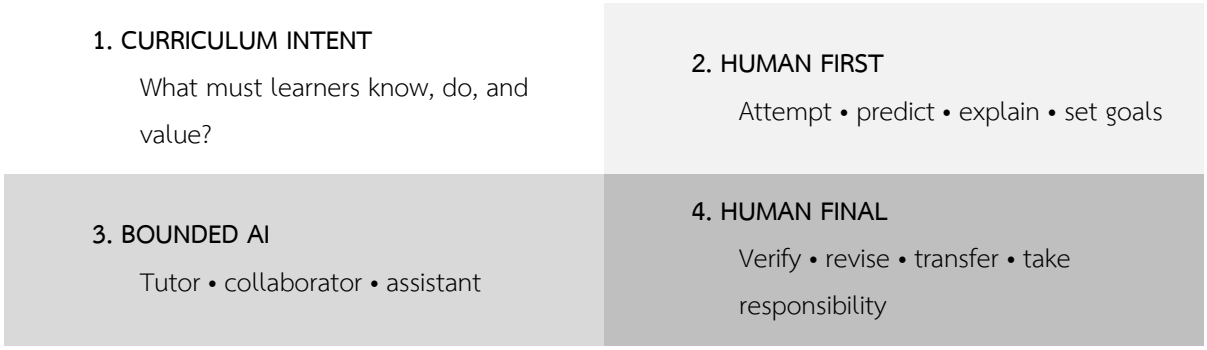


Figure 2. *Author-constructed human-centered framework for GenAI curriculum design in basic education.*

Implications and Recommendations

The recommendations in Table 5 distinguish actions by actor and indicate the confidence of the supporting evidence. “High” confidence here denotes strong convergence in risk or governance guidance, not necessarily a large causal effect on achievement. “Moderate” denotes convergent but mainly short-term or context-dependent evidence. “Emerging” denotes a plausible proposition requiring stronger school-level testing.

Table 5. *Actionable recommendations by actor*

Actor and priority action	Rationale	Confidence
Teachers — Use a human-first attempt, bounded AI assistance, and a human-final explanation or product	Preserves productive struggle and makes learner contribution visible	Moderate
Teachers — Require source verification, correction logs, and proportional disclosure	Addresses fluent error, provenance, and accountability	High for risk control; emerging for learning effect
Teachers — Include AI-free retrieval, transfer, or oral-defense checkpoints	Distinguishes assisted performance from acquired competence	Moderate
School leaders — Approve tools	School-age learners require	High

Actor and priority action	Rationale	Confidence
through age, privacy, accessibility, language, and bias review; provide a non-AI alternative	stronger protection than adult consumers	
School leaders — Allocate time for subject-specific lesson study and shared evaluation of outputs	Teacher judgment is central to effective and safe implementation	Moderate
Curriculum agencies — Embed conceptual, practical, evaluative, and ethical AI literacy across subjects and grade bands	Prompting alone is an incomplete and unstable curriculum	Moderate-to-high framework convergence
Curriculum agencies — Define which competencies must remain independently performable	Prevents silent erosion of core knowledge and skills	Moderate
Policymakers and procurers — Require Thai-language, cultural, rural-access, security, retention, and model-update evidence	Global model performance does not guarantee local fitness	High for governance need; emerging for optimal metric
Researchers — Report model/version, prompt design, teacher role, comparison condition, and independent outcomes	Improves replicability and separates tool effects from activity design	High
Researchers — Conduct longitudinal, multi-site, and under-resourced-context studies with unaided transfer measures	Current evidence is short-term and contextually narrow	High as a research priority

Implementation should be staged. A school can first classify uses by risk and pedagogical purpose, then pilot low-risk activities with explicit success and stop criteria. Success criteria should include independent learning, equity, teacher workload, privacy incidents, and student agency—not usage volume. Results should be reviewed before expansion, and model updates should trigger renewed validation. This process turns adoption into accountable educational improvement rather than irreversible technological momentum.

Limitations and Future Research

This review has four limitations. First, it is a critical integrative review, not a new exhaustive systematic search; the purposive corpus may omit relevant studies, and no inter-rater reliability statistic is available. Second, English-language and publication bias likely shape the international evidence, while Thai and other underrepresented-language research remains limited. Third, included reviews overlap substantially, so their

study counts cannot be pooled and apparent convergence may partly reflect repeated primary studies. Fourth, GenAI systems change faster than publication cycles: model versions, safety controls, prices, and data policies may make specific findings obsolete.

These limits reduce confidence in universal effect claims but do not remove the need for action. Future research should prioritize durable and independent learning, developmental differences, students with disabilities, regional and minority languages, rural infrastructure, teacher workload, and child-data governance. Multi-site Thai studies should preregister outcomes, preserve non-AI comparison conditions, document model versions and teacher mediation, and include delayed transfer tasks. Researchers should also test the 5E and human-centered frameworks proposed here rather than treating them as established interventions.

■ Conclusion

GenAI can enrich basic education when it increases feedback, alternatives, access, and reflection without displacing the cognitive and ethical work that students must learn to perform. The current evidence supports structured, teacher-guided use more strongly than unrestricted access. It supports short-term performance and engagement claims more strongly than durable learning, transfer, or system-wide transformation. The most consequential risk is therefore not only inaccurate output; it is mistaking an improved AI-assisted product for improved human capability.

A defensible curriculum integrates four dimensions of AI literacy—conceptual understanding, practical interaction, critical evaluation, and ethical agency—while retaining independent checkpoints. Teachers require subject-specific competence and authority, learners require age-appropriate protection and meaningful choice, and institutions require language, culture, equity, and data-governance safeguards. In Thailand, locally validated Thai-language performance, cultural responsiveness, rural feasibility, and teacher readiness are conditions of responsible adoption. The educational aim should not be to make schools maximally automated. It should be to develop learners who can think, create, collaborate, and act responsibly with AI while remaining capable without it.

■ References

Alfarwan, A. (2025). Generative AI use in K-12 education: A systematic review. *Frontiers in Education, 10*, 1647573. https://doi.org/10.3389/feduc.2025.1647573

Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology, 15*(4), ep448. https://doi.org/10.30935/cedtech/13417

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology, 15*(3), ep430. https://doi.org/10.30935/cedtech/13176

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. BSCS.

Casal-Otero, L., Catala, 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*, 29. https://doi.org/10.1186/s40594-023-00418-7

Cheah, Y. H., & Kim, J. (2025). STEM teachers' perceptions, familiarity, and support needs for integrating generative artificial intelligence in K-12 education. *School Science and Mathematics*. Advance online publication. https://doi.org/10.1111/ssm.18334

Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development, 72*, 2401–2416. https://doi.org/10.1007/s11423-024-10366-w

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Dewey, J. (1916). *Democracy and education*. Macmillan.

Han, A., Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., & Peppler, K. (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, pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642438

Hays, L., Jurkowski, O., & Sims, S. K. (2024). ChatGPT in K-12 education. *TechTrends, 68*, 281–294. https://doi.org/10.1007/s11528-023-00924-z

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

International Telecommunication Union. (2022). *E-learning in Thailand: Mapping the digital divide*. https://www.itu.int/hub/publication/d-pref-them-28-2022/

Jauhiainen, J. S., & Guerra, A. G. (2023). Generative AI and ChatGPT in school children's education: Evidence from a school lesson. *Sustainability, 15*(18), 14025. https://doi.org/10.3390/su151814025

Kim, D., Lee, S., Kim, Y., Rutherford, A., & Park, C. (2025). Representing the under-represented: Cultural and core capability benchmarks for developing Thai large language models. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 4114–4129). Association for Computational Linguistics. https://aclanthology.org/2025.coling-main.278/

Kong, S. C., & Yang, Y. (2024). A human-centered learning and teaching framework using generative artificial intelligence for self-regulated learning development through domain knowledge learning in K–12 settings. *IEEE Transactions on Learning Technologies, 17*, 1588–1599. https://doi.org/10.1109/TLT.2024.3392830

Lin, X., & Tan, H. (2025). A systematic review of generative AI in K-12: Mapping goals, activities, roles, and outcomes via the 3P model. *Systems, 13*(10), 840. https://doi.org/10.3390/systems13100840

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Ma, M., Ng, D. T. K., Liu, Z., & Wong, G. K. W. (2025). Fostering responsible AI literacy: A systematic review of K-12 AI ethics education. *Computers and Education: Artificial Intelligence, 8*, 100422. https://doi.org/10.1016/j.caeai.2025.100422

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Miao, F., & Shiohira, K. (2024). *AI competency framework for students*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391105

Min, T., Lee, B., & Jho, H. (2025). Integrating generative artificial intelligence in the design of scientific inquiry for middle school students. *Education and Information Technologies, 30*, 15329–15360. https://doi.org/10.1007/s10639-025-13410-1

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Nantaburom, S. (2026). Perspectives of secondary school students toward using generative AI as a summative assessment tool of social studies courses in Thailand. *Discover Education, 5*, 549. https://doi.org/10.1007/s44217-026-01565-z

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*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Organisation for Economic Co-operation and Development. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing. https://doi.org/10.1787/c74f03de-en

Organisation for Economic Co-operation and Development. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. https://doi.org/10.1787/062a7394-en

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

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*. Advance online publication. https://doi.org/10.1080/15391523.2025.2455057

Su, J., & Yang, W. (2024). Powerful or mediocre? Kindergarten teachers' perspectives on using ChatGPT in early childhood education. *Interactive Learning Environments, 32*(10), 6496–6508. https://doi.org/10.1080/10494820.2023.2266490

Tang, K. S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning and assessment with generative AI in secondary education: A proof of concept. *Pedagogies: An International Journal, 19*(3), 493–503. https://doi.org/10.1080/1554480X.2024.2379774

Tongchai, A., & Malakul, S. (2026). Thai teachers' perceptions of integrating generative AI in K-12 education: Opportunities and challenges. *TechTrends, 70*, 93–105. https://doi.org/10.1007/s11528-025-01128-3

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review, 15*(4), 404–428. https://doi.org/10.1177/1534484316671606

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

UNESCO Global Education Monitoring Report Team. (2023). *Global education monitoring report 2023: Technology in education—A tool on whose terms?* UNESCO. https://www.unesco.org/gem-report/en/technology

UNICEF. (2025). *Guidance on AI and children* (Version 3.0). UNICEF Innocenti. https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

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

Wang, N., Wang, X., & Su, Y.-S. (2024). Critical analysis of the technological affordances, challenges and future directions of generative AI in education: A systematic review. *Asia Pacific Journal of Education, 44*(1), 139–155. https://doi.org/10.1080/02188791.2024.2305156

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Yang, T. C., Hsu, Y. C., & Wu, J. Y. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: Evidence from a quasi-experimental research. *Interactive Learning Environments, 33*(6), 3726–3743. https://doi.org/10.1080/10494820.2025.2450659

Yusuf, A., Pervin, N., Román-González, M., & Noor, N. M. (2024). Generative AI in education and research: A systematic mapping review. *Review of Education, 12*(2), e3489. https://doi.org/10.1002/rev3.3489

Zhang, L., Deng, X., & Shadiev, R. (2026). A meta-review of generative AI in education: Synthesizing findings from systematic reviews. *Journal of Educational Computing Research, 64*(4), 1068–1092. https://doi.org/10.1177/07356331261419689

Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. *European Journal of Education, 59*(2), e12599. https://doi.org/10.1111/ejed.12599

Zhang, T., Lai, Y. C., & Yu, P. L. H. (2026). Generative artificial intelligence in K-12 education: A systematic review. *Research and Practice in Technology Enhanced Learning, 21*, 34. https://doi.org/10.58459/rptel.2026.21034