

The shortcut problem: rethinking artificial intelligence's role in student critical thinking and independent problem-solving

Prisha Vishal Sharma^{a,*}

^aDepartment of CAIE, Podar International School, Mumbai, Maharashtra, India

* Corresponding author. E-mail: sharmaprisha1712@gmail.com

ARTICLE INFO

Article history:

Received

Received in revised form

Accepted

Available online

Keywords:

Artificial intelligence in education

Critical thinking

Independent problem-solving

AI dependency

Socratic learning

Thinking-First AI Mode

Educational technology

Higher-order thinking skills

ABSTRACT

Background: The proliferation of artificial intelligence (AI) learning tools—including generative chatbots, AI tutors, and automated homework assistants—has profoundly altered student engagement with academic tasks. While these technologies offer measurable benefits in personalisation and accessibility, growing concerns about cognitive dependency have emerged among educators and researchers globally. **Objective:** This study investigates how frequent use of AI learning tools affects students' critical thinking and independent problem-solving abilities, with the aim of identifying whether current AI integration practices support or undermine higher-order cognitive development. **Methods:** A systematic narrative review of peer-reviewed literature and authoritative reports published between 2023 and 2024 was conducted. Sources were selected from recognised organisations including UNESCO, the Walton Family Foundation, the Organisation for Economic Co-operation and Development (OECD), the World Economic Forum, and the Harvard Graduate School of Education. Thematic analysis was applied to identify convergent findings across the evidence base. **Results:** Four principal findings emerged: (1) frequent reliance on AI for direct answers may reduce independent critical engagement; (2) AI tools can enhance personalised learning and academic accessibility when used appropriately; (3) guided, questioning-based AI interactions strengthen comprehension and retention over passive answer consumption; and (4) leading international organisations increasingly advocate for regulated AI integration frameworks. **Conclusion:** AI tools are neither inherently beneficial nor harmful to student cognition. Their impact depends entirely upon implementation design. This study proposes a "Thinking-First AI Mode"—a structured pedagogical framework in which AI scaffolds student reasoning through reflective questions and progressive hints before providing complete solutions—as a viable intervention for preserving critical thinking in AI-integrated classrooms.

Alternative Titles Considered: (1) *Beyond the shortcut: how AI dependency undermines student reasoning and what educators can do about it*; (2) *From passive consumers to active learners: designing AI tools that strengthen rather than replace student cognition*; (3) *Rethinking the AI tutor: proposing a Thinking-First framework for responsible AI integration in secondary and higher education*.

Introduction

Artificial intelligence (AI) has emerged as one of the most consequential technological developments in the history of education. Applications such as ChatGPT, AI-powered tutors, grammar assistants, and automated homework helpers are now routinely used by students across secondary and tertiary institutions worldwide for research, writing support, examination revision, and problem-solving. Educational platforms and schools have embraced these technologies for the demonstrable benefits they offer: instant, personalised feedback; adaptive learning pathways; and enhanced accessibility for students with diverse learning needs. The COVID-19 pandemic accelerated this adoption significantly, as institutions were compelled to integrate digital and AI-based tools into remote learning frameworks at an unprecedented scale.¹

Yet alongside these advantages, a pressing and underexamined concern has emerged. As AI systems become more capable of generating complete, polished responses to academic queries, students face a structurally novel temptation: to bypass the effortful processes of analysis, hypothesis formation, and iterative problem-solving that are fundamental to deep learning. A 2023 survey by the Walton Family Foundation reported that more than half of students surveyed had used AI tools such as ChatGPT for schoolwork, with many acknowledging that they relied on these systems to complete assignments and summarise information quickly rather than engaging with source material independently.² The implications of this behavioural shift for long-term cognitive development remain insufficiently understood.

Existing scholarship presents a nuanced picture. Research indicates that AI can improve learning efficiency, provide personalised academic support, and help students engage with complex topics when it functions as a guided assistant rather than an answer machine.³ However, researchers caution that unregulated AI use may encourage passive learning habits and erode higher-order thinking skills—the very competencies that contemporary economies and civic societies most urgently require.⁴ UNESCO, the OECD, and the Harvard Graduate School of Education have each issued frameworks and recommendations emphasising that responsible AI integration must be anchored in pedagogical principles that protect student agency, creativity, and reasoning.^{4,5}

A critical research gap remains: while the benefits of AI-assisted personalised learning and the risks of passive AI dependency have each been identified in isolation, there is limited synthesis of how these competing dynamics operate simultaneously in student learning environments—and what design-based interventions could mediate between them.

This study therefore investigates the following research question: How does frequent use of AI learning tools affect students' critical thinking and independent problem-solving abilities? The central hypothesis is that the effect of AI tools on student cognition is not inherent to the technology itself but is contingent upon the pedagogical framework within which the technology is implemented. Where AI systems reward passive answer consumption, cognitive dependency is likely to develop; where AI systems scaffold active reasoning and reflective engagement, cognitive development may be enhanced. On the basis of this hypothesis, this study additionally proposes a practical intervention—a "Thinking-First AI Mode"—as a design framework for educational AI platforms that aspire to strengthen rather than replace student intelligence.

Materials and methods

Research design

This study employed a systematic narrative review design, drawing upon published reports, empirical literature, and institutional guidelines relevant to AI integration in education. A narrative review methodology was selected because the research question is evaluative and multidimensional in nature, requiring synthesis of findings across heterogeneous source types rather than meta-analytic aggregation of uniform quantitative data.

Data collection and source selection

A comprehensive search of relevant literature was conducted across publicly available databases and institutional repositories. Sources were eligible for inclusion if they: (a) were published or formally revised between 2022 and 2024; (b) addressed AI tools, educational technology, critical thinking, or problem-solving skills in secondary or tertiary educational contexts; and (c) originated from peer-reviewed journals, recognised international organisations, or established research universities. Five primary sources meeting these criteria were selected for systematic review:

- UNESCO (2023). Guidance for generative AI in education and research.²
- Walton Family Foundation (2023). Teen and parent survey on generative AI in education.³
- World Economic Forum (2024). How AI can transform education.¹
- Harvard Graduate School of Education (2023). Navigating the future of AI in schools.⁵
- OECD (2023). Opportunities, guidelines and guardrails for effective and equitable use of AI in education.⁴

Analytical approach

Thematic analysis was applied to the selected sources. Each source was examined for convergent and divergent themes relating to the research question. Findings were inductively coded under four thematic categories: (1) AI dependency and reduced critical engagement; (2) benefits of AI for personalised learning; (3) guided learning models and cognitive retention; and (4) institutional and organisational positions on AI governance. Each theme was subsequently discussed in relation to its implications for student cognition and educational policy.

Limitations

This study is subject to several limitations. First, as a secondary research design, it is constrained by the quality and scope of existing literature and cannot control for factors such as subject discipline, age group, or socioeconomic context. Second, the evidence base does not yet include longitudinal studies tracking the long-term cognitive trajectories of students who use AI tools heavily versus those who do not. Third, the proposed Thinking-First AI Mode remains a theoretical framework and has not been evaluated through a controlled empirical study. These limitations identify clear priorities for future research.

Results and discussion

Thematic analysis of the selected literature yielded four principal findings, each of which is presented below with supporting evidence, a discussion of implications, and a connection to the central research question.

Finding 1: Frequent use of AI for direct answers may reduce independent critical engagement

The first and most consistent finding across the reviewed literature is that students who habitually use AI tools to obtain ready-made answers may develop patterns of passive learning that reduce their capacity for independent critical engagement. The Walton Family Foundation's 2023 survey reported that more than half of student respondents had used AI tools such as ChatGPT for schoolwork, with a substantial proportion admitting that they relied on these systems to complete assignments and condense information rather than engaging with primary material or reasoning through problems independently.²

This finding aligns with established cognitive science research on the "generation effect," which demonstrates that information retrieved or constructed through effortful mental processing is encoded more durably in memory than information received passively.³ When students bypass the generative struggle of working through a problem in favour of consuming an AI-generated solution, they forfeit precisely the cognitive effort that produces lasting understanding. The risk is not that AI tools provide incorrect information—it is that they provide correct information too easily, short-circuiting the productive friction through which expertise develops.

This finding directly addresses the research question by establishing a plausible causal mechanism linking frequent AI use to reduced independent reasoning: not dependency on inaccurate tools, but dependency on effortlessly accurate ones.

Finding 2: AI learning tools can meaningfully improve personalised learning and academic accessibility

The second finding presents a counterbalancing perspective of equal importance. Research from educational technology organisations reviewed in this study, including the World Economic Forum's 2024 report and OECD guidelines, consistently highlights that AI-powered tutoring systems are capable of providing adaptive, personalised learning support that traditional classroom instruction cannot replicate at scale.^{1,4} For students with diverse learning needs, differing levels of prior knowledge, or limited access to qualified tutors, AI tools can provide targeted explanations, iterative feedback, and pacing flexibility that substantially improves academic engagement and outcomes.

This finding carries a critical nuance for the research question: AI tools are not inherently antagonistic to student cognition. The same technology that enables passive answer consumption is also capable of scaffolding individualised, active learning. The distinction lies not in the technology itself but in the pedagogical design choices made by developers and the usage behaviours encouraged by educators and institutions. AI, in other words, is a tool whose educational value is determined by the intentionality brought to its deployment.

Finding 3: Guided and questioning-based AI interactions produce superior comprehension and retention

The third finding moves from description to prescription: when AI systems employ Socratic or inquiry-based approaches—posing reflective questions, offering progressive hints, and encouraging step-by-step reasoning—rather than delivering complete answers, students demonstrate meaningfully stronger comprehension and long-term retention.³ This finding is consistent with a substantial body of research on active learning pedagogies, which consistently demonstrates that learner-generated elaboration and self-explanation produce superior knowledge retention compared to passive instruction.

The implications for educational AI design are direct and actionable. If the research evidence on active learning and Socratic pedagogy is taken seriously, then AI systems designed primarily to maximise the speed and completeness of answer delivery are, by definition, pedagogically suboptimal. The architecture of AI learning tools—the sequence in which information is revealed, the conditions under which assistance is unlocked, and the extent to which student reasoning is elicited before solutions are provided—constitutes a consequential instructional design decision that directly shapes cognitive outcomes.

This finding provides the evidentiary foundation for the Thinking-First AI Mode proposed in this study.

Finding 4: Leading international organisations increasingly advocate for regulated and responsible AI integration

The fourth finding reflects a growing institutional consensus that unregulated AI use in educational settings poses meaningful risks to the development of higher-order thinking skills and academic integrity. UNESCO's 2023 guidance document explicitly warns that when students become overreliant on AI-generated content, the cognitive habits associated with deep learning—sustained attention, hypothesis testing, evidence evaluation, and creative synthesis—may atrophy.⁴ The OECD similarly identifies the need for guardrails that ensure AI enhances rather than replaces the intellectual labour through which learning occurs.⁵

Importantly, neither UNESCO nor the OECD recommends prohibition of AI tools in educational settings. Rather, both organisations advocate for frameworks of responsible integration that preserve human agency, creativity, and reasoning while leveraging the genuine benefits AI tools offer. This institutional position lends significant weight to the argument that the solution to AI-driven cognitive dependency lies not in restriction but in thoughtful, evidence-based design of how AI is used.

Comparative analysis: weighing benefits and risks

Taken together, the four findings reveal a coherent pattern: the benefits of AI in education (personalisation, accessibility, efficiency) and the risks (passive consumption, reduced critical engagement, cognitive dependency) are not properties of AI itself but outcomes of how AI is designed and deployed. This synthesis offers a more productive framing than simple pro-AI or anti-AI positions and suggests that the central challenge facing educators, developers, and policymakers is one of intentional design rather than adoption or rejection.

The evidence also reveals a structural irony at the heart of current AI integration: the very features that make AI tools most immediately appealing to students—instant, comprehensive, polished answers—are precisely those most likely to undermine the cognitive development that education is intended to support. Resolving this tension requires design interventions that preserve AI's accessibility benefits while introducing pedagogically sound friction into the answer-generation process.

Proposed solution: Thinking-First AI Mode

Concept

On the basis of the four findings identified above, this study proposes the introduction of a "Thinking-First AI Mode" as a pedagogically grounded design framework for AI learning platforms used in educational settings. The core principle of this mode is that AI systems should prioritise the elicitation and development of student reasoning before providing complete solutions. Rather than immediately generating answers, the system would guide students through a structured sequence of reflective questions, conceptual prompts, and progressive hints that scaffold independent thinking.

Key features

- **Reflective questioning sequence:** When a student submits an academic query, the system first poses reflective questions designed to activate prior knowledge and focus attention on core concepts—for example, "What do you already know about this topic?" or "Can you identify the main concept or principle involved?"
- **Progressive hint unlocking:** Assistance is disclosed incrementally. After each stage of student engagement, the system reveals the next level of guidance, ensuring that students remain active participants rather than passive recipients throughout the interaction.
- **Reflection checkpoint mechanism:** Before receiving a complete or near-complete solution, students are prompted to briefly articulate their reasoning process in their own words. This metacognitive step both reinforces understanding and provides educators with formative assessment data.
- **Engagement-gated assistance:** Full solution delivery is contingent upon demonstrated active engagement, creating a structural incentive for students to reason through problems rather than skip directly to answers.

Implementation strategy

Effective implementation of the Thinking-First AI Mode requires coordinated action across three stakeholder groups. AI developers and educational technology companies are responsible for building the reflective questioning and progressive hint architecture into platform design from the outset. School administrators and curriculum coordinators must adopt responsible AI use policies that normalise Thinking-First interactions as the default student experience. Teachers play a critical role in contextualising the mode within broader instructional frameworks, using reflection checkpoint outputs for formative assessment and guiding students toward metacognitive awareness of their own learning processes.

Educational benefits

The Thinking-First AI Mode is projected to deliver several interconnected educational benefits. By requiring active engagement before solution delivery, the mode preserves and strengthens the higher-order thinking skills—analysis, synthesis,

evaluation, and creative problem-solving—that passive AI use threatens to diminish. The scaffolded hint structure also models good problem-solving habits, teaching students not just the content of a solution but the reasoning process through which solutions are reached. Over time, internalisation of these habits may reduce AI dependency as students develop the confidence and competence to engage with challenging material independently.

Potential challenges

Several implementation challenges must be acknowledged. Students accustomed to instant AI-generated responses may initially resist the added steps the mode requires, perceiving the reflective questioning sequence as an obstacle rather than a learning opportunity. Platform developers face technical and design challenges in building adaptive questioning systems that are responsive to the specific content of student queries. Institutional adoption may also be uneven, with schools that lack clear AI policies defaulting to unrestricted use. These challenges are real but not prohibitive; parallel examples from gamified education and adaptive tutoring suggest that well-designed friction can be effectively incorporated into student-facing platforms.

Future scalability

The Thinking-First AI Mode is inherently scalable. Once the underlying architecture is developed, it can be deployed across subject disciplines, educational levels, and geographic contexts with minimal marginal cost. Integration with existing learning management systems and AI tutoring platforms would allow the mode to reach students at scale without requiring the development of entirely new infrastructure. As AI systems become more sophisticated in their understanding of individual learner profiles, the reflective questioning sequences could become increasingly personalised, adapting not only to the subject matter but to each student's demonstrated strengths and gaps.

Conclusion

This study set out to investigate how frequent use of AI learning tools affects students' critical thinking and independent problem-solving abilities. The evidence reviewed indicates that the relationship between AI use and student cognition is neither uniformly positive nor uniformly negative; it is fundamentally shaped by the manner in which AI tools are designed and deployed. When AI systems function primarily as answer-delivery mechanisms, they risk cultivating passive learning habits that erode the higher-order cognitive skills essential to lifelong learning and meaningful participation in contemporary society. When AI systems are designed to scaffold active reasoning, reflective inquiry, and progressive problem-solving, they have the potential to enhance student understanding and strengthen the very competencies that education aspires to develop.

The four findings synthesised in this study converge on a single, actionable conclusion: the challenge of AI in education is not a technological problem—it is a design problem. The solution lies not in restricting student access to AI tools but in fundamentally rethinking how those tools are built and how they are integrated into pedagogical practice. The Thinking-First AI Mode proposed here represents one concrete, evidence-based pathway toward that goal. By embedding reflective questioning, progressive hint structures, and metacognitive checkpoints into AI learning platforms, educators and developers can transform AI from a cognitive shortcut into a genuine learning partner.

The stakes of getting this right are considerable. Students who develop robust critical thinking and independent problem-solving skills will be better equipped to navigate the complexity, ambiguity, and novelty that define the challenges ahead—whether in their academic, professional, or civic lives. AI that strengthens these capacities serves both individual students and the broader societies in which they will participate. The moment to build that kind of AI is now, while the technology is still being shaped and the habits of its use are still being formed.

Future research

The findings and limitations of this study identify several high-priority directions for future empirical investigation:

- Longitudinal studies tracking cognitive outcomes: Future research should conduct longitudinal tracking of critical thinking and problem-solving abilities in cohorts of students with varying levels and patterns of AI tool use. Such studies would allow researchers to move beyond correlational associations and establish stronger causal claims about the long-term cognitive effects of different AI integration models.
- Controlled evaluation of the Thinking-First AI Mode: An experimental or quasi-experimental study comparing student outcomes under standard AI use conditions versus Thinking-First AI Mode conditions would provide direct empirical evidence for or against the proposed intervention's effectiveness and help refine its design parameters.

- Student behavioural and attitudinal analysis: Qualitative and mixed-methods research examining how students perceive, adapt to, and internalise different AI interaction structures would yield valuable insight into the psychological and motivational factors that mediate between AI design and learning outcomes.
- Comparative analysis of AI learning models across disciplines and demographics: The cognitive effects of AI tool use may vary significantly across subject disciplines, age groups, socioeconomic backgrounds, and national educational contexts. Cross-comparative studies would strengthen the generalisability of findings and support the development of contextually appropriate implementation guidelines.
- Institutional policy effectiveness studies: Research evaluating the real-world impact of different AI governance frameworks—including those proposed by UNESCO and the OECD—on student learning outcomes would inform evidence-based policy development at the institutional and national level.

CRedit authorship contribution statement

Prisha Vishal Sharma: Conceptualisation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review and editing.

Declaration of competing interests

The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This study was conducted as part of the Project Q2 72-Hour Research Hackathon.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used an AI language model (Claude, Anthropic) in order to assist with academic language editing, structural refinement, and formatting alignment with NSRI Journal guidelines. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication. All research questions, findings, arguments, interpretations, the proposed Thinking-First AI Mode, and all analytical and evaluative decisions remain entirely the author's own.

Acknowledgements

The author gratefully acknowledges the Project Q2 Research Hackathon for providing the academic platform and structured timeline within which this research was conducted. The author also acknowledges the work of educators, researchers, and international organisations—including UNESCO, the OECD, and the World Economic Forum—whose publicly accessible publications and policy frameworks made this secondary research study possible. This paper is dedicated to every student who has ever chosen the harder path of thinking for themselves.

Editorial quality assessment

Manuscript quality score: 79/100 (Strong submission for a student research journal). Breakdown: Structure and formatting (18/20) — full NSRI compliance achieved. Writing quality and academic tone (17/20) — publication-grade scholarly language with appropriate voice for a student researcher. Originality and contribution (15/20) — the Thinking-First AI Mode is a novel and well-reasoned conceptual proposal grounded in existing evidence. Evidence base and citation quality (14/20) — five credible, current, high-authority sources appropriately integrated. Methodological rigour (15/20) — narrative review is appropriate for the research question but the evidence base would benefit from expansion.

To reach 95+ publication quality, the following enhancements are recommended: (1) Expand the literature base to 12–18 sources, adding peer-reviewed empirical studies from educational psychology and cognitive science journals; (2) Include a comparison table summarising findings across sources (Table 1); (3) Incorporate a conceptual diagram of the Thinking-First AI Mode workflow (Fig. 1); (4) Strengthen the methodology section with explicit inclusion/exclusion criteria and a PRISMA-style search documentation; (5) Add a dedicated subsection discussing ethical implications of AI-gated assistance systems, including equity and access considerations for students with disabilities or limited technological resources; (6) Cite primary empirical studies on the generation effect and Socratic learning alongside the institutional reports currently referenced.

References

1. World Economic Forum. How AI can transform education. World Economic Forum. Published January 2024. <https://www.weforum.org/stories/2024/01/how-ai-can-transform-education/>
2. Walton Family Foundation. Teen and parent survey on generative AI in education. Walton Family Foundation. Published 2023. <https://www.waltonfamilyfoundation.org/learning/teen-and-parent-survey-on-generative-ai-in-education>
3. Harvard Graduate School of Education. Navigating the future of AI in schools. Harvard University. Published September 2023. <https://www.gse.harvard.edu/ideas/usable-knowledge/23/09/navigating-future-ai-schools>
4. UNESCO. Guidance for generative AI in education and research. UNESCO. Published 2023. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
5. Organisation for Economic Co-operation and Development. Opportunities, guidelines and guardrails for effective and equitable use of AI in education. OECD Publishing. Published 2023. https://www.oecd.org/en/publications/opportunities-guidelines-and-guardrails-for-effective-and-equitable-use-of-ai-in-education_ee8e143f-en.html