Nova Science Press

Journal of Psychology & Education

Vol. 1, No. 9 (2026)

When Answers Come Before Questions: Does Generative Artificial Intelligence Weaken Students' Independent Thinking?

An Educational Argument for Protecting Cognitive Agency

Xia Lang[1] *

1 INTI International University

*Corresponding author: Xia Lang; email: i24028557@student.newinti.edu.my

Journal of Psychology & Education • Vol. 1, No. 9 (2026)

DOI:https://doi.org/10.66581/4gskcp89

Received: 16 August 2026 • Accepted: 18 October 2026 • Published: 31 October 2026

CITATION

Xia, L. (2026). When Answers Come Before Questions: Does Generative Artificial Intelligence Weaken Students' Independent Thinking?. Journal of Psychology & Education, 1(9), 41-58. https://doi.org/10.66581/4gskcp89

COPYRIGHT

© 2026 Lang, X. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/). Use, distribution, and reproduction in any medium are permitted, provided the original author(s) and the copyright owner(s) are credited and the original publication in this journal is cited in accordance with accepted academic practice.

Abstract

The rapid adoption of generative artificial intelligence (GenAI) in education has renewed a fundamental question: does convenient access to machine-generated explanations, arguments, and solutions weaken students' willingness and capacity to think independently? The question cannot be answered by treating AI use as a single behavior or equating fluent academic products with genuine learning. This conceptual and argumentative article examines the conditions under which GenAI may displace, rather than support, essential cognitive activity. Drawing on cognitive offloading, self-regulated learning, cognitive load theory, the ICAP framework, and emerging experimental and review evidence, it distinguishes immediate task performance from durable competence and behavioral compliance from intellectual agency. Three risks are emphasized: premature delegation of problem formulation and reasoning, uncritical acceptance of plausible outputs, and reduced practice in monitoring and correcting one's own understanding. Equally, structured AI tutoring, dialogic challenge, formative feedback, and accessibility support can extend opportunities for deeper learning. The article proposes a pedagogical principle of protected cognitive agency: students should retain responsibility for defining questions, constructing initial explanations, evaluating evidence, and demonstrating independent transfer, even when AI assists with peripheral operations. A four-phase instructional sequence—attempt, consult, contest, and reconstruct—is offered as a practical heuristic rather than an empirically validated intervention. Implications for assessment, teacher practice, equity, and longitudinal research are discussed. The central conclusion is conditional: GenAI does not inevitably diminish independent thinking, but educational systems can create that outcome when they reward polished answers without protecting the mental work through which understanding is formed.

Keywords: generative artificial intelligence; independent thinking; cognitive agency; critical thinking; cognitive offloading; self-regulated learning; higher education; educational assessment

Article type: Conceptual and argumentative article

I. INTRODUCTION: THE EDUCATIONAL PROBLEM BEHIND THE CONVENIENT ANSWER

For a student confronted with an unfamiliar concept, a blank page, or a complex mathematical problem, generative artificial intelligence offers an apparently ideal response: an explanation arrives immediately, a structure appears without hesitation, and a persuasive answer can be produced in seconds. This convenience is not merely another improvement in access to information. A search engine normally offers material that students must interpret and connect; an AI chatbot can present the interpretation, connection, and final expression together. It can therefore participate in the very intellectual operations that schooling is intended to cultivate.

The resulting controversy is often framed as a choice between technological progress and intellectual decline. Optimistic accounts emphasize personalized tutoring, faster feedback, expanded access, and the democratization of academic assistance. Pessimistic accounts warn that students will cease to question, tolerate uncertainty, or persist through demanding intellectual tasks. Both positions identify genuine possibilities, but neither adequately explains why the same technology can support serious inquiry in one classroom and superficial completion in another. The more consequential issue is the organization of cognitive labor: what does the student still have to think through, and what has the system already thought through on the student's behalf?

This distinction matters because education has two different outputs that frequently look alike. One is a completed task: an essay, worked solution, presentation, or research summary. The other is a learner who has acquired knowledge, judgment, and habits that can be applied to a new task without external assistance. GenAI can improve the former without necessarily improving the latter. Indeed, the gap between visible product quality and invisible cognitive development may be one of the defining educational challenges of the current period. A professionally written response is not proof that its apparent author can explain, defend, or reconstruct the reasoning behind it.

There is already empirical reason to resist simplistic conclusions. In a field experiment with high-school mathematics students, Bastani et al. (2025) found that access to a relatively unrestricted AI assistant improved performance during practice but was associated with poorer subsequent performance when assistance was removed; the negative learning effect was largely mitigated by a more carefully constrained tutoring system. Conversely, Kestin et al. (2025) reported learning advantages from an AI tutor intentionally designed around established instructional principles. These studies differ in setting, design, and comparison, so neither should be treated as a universal verdict. Together, however, they suggest that the educational consequences of AI depend not simply on access but on the kind of interaction that access makes likely.

This article argues that the relevant danger is not the disappearance of thought but the weakening of **cognitive agency**: students' capacity and disposition to initiate inquiry, choose a strategy, evaluate reasons, monitor misunderstanding, and take responsibility for conclusions. The argument is developed in five steps. First, independent thinking is defined more precisely than the popular phrase “thinking for oneself.” Second, the cognitive mechanisms through which assistance can become substitution are examined. Third, current evidence is assessed without confusing short-term productivity with long-term learning. Fourth, a conditional framework is presented for distinguishing productive from unproductive AI use. Finally, implications are drawn for classroom design, assessment, educational equity, and future research. The objective is neither to prohibit AI nor to celebrate its novelty, but to identify which intellectual activities education must preserve in an AI-rich environment.

II. WHAT DOES IT MEAN FOR STUDENTS TO THINK INDEPENDENTLY?

Independent thinking should not be confused with solitary thinking. Students develop reasoning through teachers, books, peers, cultural tools, and disciplinary communities. A student who consults a textbook or debates an idea with a classmate is not thereby less independent. Independence refers instead to the learner's control over the purposes, justification, and revision of an intellectual judgment. It is possible to think independently while using assistance, just as it is possible to behave passively while working alone.

At least four dimensions deserve separate consideration. **Initiative** is the readiness to recognize a problem, formulate a question, or pursue a discrepancy rather than merely wait for instructions. **Generative reasoning** is the ability to produce tentative explanations, interpretations, hypotheses, or solution paths from one's current knowledge. **Epistemic evaluation** involves judging the reliability of claims, the relevance of evidence, and the strength of alternative explanations. **Metacognitive regulation** concerns planning, monitoring comprehension, identifying errors, and adjusting strategies. None is equivalent to producing a polished final document, and each can be obscured when an AI system delivers that document.

The ICAP framework provides a useful lens. Chi and Wylie (2014) distinguish passive, active, constructive, and interactive forms of engagement, with stronger learning expected when learners generate new inferences and build ideas through meaningful interaction rather than merely receive information. AI use can occupy any of these modes. Copying an explanation is predominantly receptive. Highlighting or editing it may involve activity without substantial conceptual construction. Asking the system to critique a student-generated hypothesis may invite constructive engagement, while sustained dialogue in which the student develops, tests, and revises an argument may approximate an interactive learning process. The presence of a chatbot is thus not itself evidence of active learning; the cognitive operations performed by the learner must be examined.

This conceptual precision also prevents unfair comparisons. Before concluding that AI has reduced independent thinking, researchers must specify the baseline: compared with independent study, ordinary internet searching, teacher tutoring, peer collaboration, or a less structured AI interface? A student using AI to translate unfamiliar terminology might become more independent in a demanding science discussion than a student excluded from such support. By contrast, one who delegates the central inference in a logic exercise may lose precisely the opportunity the exercise was designed to provide. The same act of assistance can have a different educational meaning according to the target competence.

III. HOW ASSISTANCE BECOMES SUBSTITUTION: COGNITIVE AND MOTIVATIONAL MECHANISMS

3.1 Cognitive offloading and the location of intellectual work

Cognitive offloading refers to the use of external resources or actions to reduce demands on internal processing (Risko & Gilbert, 2016). Writing down a calculation, using a calendar, or consulting a reference work can be highly adaptive. Human cognition has never operated without tools, and efficient external support can release attention for demanding decisions. The educational problem begins when the delegated activity is not a dispensable burden but the very operation students need to learn.

Consider two uses of AI during an argument-writing assignment. A student who has already developed a position may ask for help identifying ambiguous transitions or improving the readability of a paragraph. Such assistance can remove peripheral obstacles while leaving argumentative decisions with the learner. Another student may prompt the system to select the thesis, provide evidence, anticipate objections, and compose the conclusion. In the second case, the essential cognitive labor has been transferred. The result may be rhetorically impressive, but the student's opportunity to practise constructing an argument has narrowed.

This distinction is compatible with Nakamura's (2026) augmentation–erosion framework, which emphasizes whether externalized operations are peripheral or constitutive of the learning goal. It also cautions against assuming that any reduction in cognitive effort is undesirable. Offloading routine formatting, spelling, or navigation may be sensible when the objective is to understand a scientific mechanism. But offloading hypothesis generation during a lesson intended to teach hypothesis generation has a fundamentally different effect. The danger is not efficiency itself; it is a mismatch between efficient task completion and the cognitive purpose of the task.

3.2 Cognitive load: less effort is not always more learning

Cognitive load theory helps clarify the apparent paradox. Learners have limited working-memory capacity, and unnecessary demands can interfere with the construction of useful knowledge structures (Sweller, 1988). An AI tutor that clarifies confusing instructions, sequences a difficult explanation, or supplies a timely hint may reduce unproductive burden and permit deeper processing. An interface that removes every meaningful challenge can, however, eliminate occasions for explanation, retrieval, discrimination, and revision.

Research on retrieval practice is relevant here. Roediger and Karpicke (2006) showed that retrieving information can improve later retention even when re-exposure feels easier. More broadly, Dunlosky et al. (2013) identify practice testing and distributed practice as effective learning techniques. AI-generated summaries can create a reassuring sense of familiarity, but recognition of a well-written explanation is not

the same as being able to reconstruct it. Students may confuse ease of comprehension while reading an answer with mastery of the process required to produce an answer independently.

3.3 Metacognitive outsourcing and false confidence

Effective independent learners do more than select correct answers: they notice uncertainty, check progress, and diagnose errors. Zimmerman's (2002) account of self-regulated learning emphasizes forethought, performance regulation, and self-reflection. If a chatbot continuously supplies plans, identifies mistakes, proposes revisions, and judges adequacy, students may have fewer reasons to exercise these regulatory functions. They can become skilled at obtaining corrections without becoming skilled at detecting the need for correction.

A related risk is miscalibrated confidence. The fluency and apparent coherence of AI-generated explanations may make understanding seem more complete than it is. Students who can follow a response may assume they could have generated it. Yet this assumption can be tested only through unaided explanation, application, or transfer. The solution is not perpetual distrust of AI, which would be exhausting and impractical, but informed judgment concerning which claims require checking and which reasoning steps a learner must be able to perform independently.

3.4 Motivation, autonomy, and intellectual persistence

Independent thinking is partly motivational. Students must decide that a question is worth pursuing and that their own effort can make a difference. Self-determination theory identifies autonomy and competence as central conditions for high-quality motivation (Ryan & Deci, 2020). GenAI can support both by offering responsive explanations and reducing the embarrassment of asking a question. It can also undermine both when learning is experienced as selecting among machine-produced answers rather than exercising authorship over an inquiry.

The mechanism is subtle. Immediate relief from uncertainty is rewarding, especially under deadlines. Repeated reliance may gradually alter students' expectations about how much personal effort is normal before seeking an external solution. This is a plausible developmental pathway, not proof that AI causes generalized laziness. It suggests, however, that the design of tasks and feedback may influence whether students interpret effort as a valuable route to competence or as an obstacle that ought always to be bypassed.

IV. WHAT DOES THE EMERGING EVIDENCE ACTUALLY ESTABLISH?

A responsible educational argument must distinguish experimental findings, research syntheses, and theoretical possibilities. The literature is expanding quickly, but studies use different tools, ages, subjects,

measures, and instructional arrangements. A claim that GenAI either improves or harms "thinking" across education would exceed what these heterogeneous outcomes can establish.

Bastani et al. (2025) provide unusually direct evidence for the distinction between supported performance and independent learning. In a field experiment involving nearly one thousand high-school mathematics students, an unrestricted GPT-based assistant produced substantial gains while it was available. Yet students in that condition scored lower on a subsequent assessment after AI access was removed than peers who had never received the tool. A more constrained AI tutor, designed to provide pedagogically useful guidance instead of simply disclosing answers, largely avoided that penalty. The lesson is specific but powerful: the same family of tools can generate very different learning outcomes depending on whether it invites answer acquisition or problem-solving.

Stadler et al. (2024) reached a related conclusion in a randomized study of 91 university students investigating a socio-scientific issue. Compared with students using conventional web search, those using a language model reported lower cognitive load but produced weaker reasoning and argumentation in their final recommendations. This finding does not demonstrate permanent cognitive decline. It does, however, challenge the assumption that faster access to synthesized information necessarily creates deeper understanding. The instructional value of cognitive ease must be judged against the quality of the reasoning that follows it.

Writing research offers another warning against substituting product quality for learning. In a randomized experiment with 117 university students comparing several forms of support, Fan et al. (2025) observed better essay-score improvement in the ChatGPT condition without corresponding significant advantages in knowledge gain or transfer. The authors also identified process patterns compatible with overdependence and weakened metacognitive engagement. These results should not be extended automatically to every writing context, but they support the view that improved artifacts and stronger learner competence are empirically separable outcomes.

Evidence on the benefits of AI is equally important. Kestin et al. (2025) evaluated a deliberately designed AI tutor in college-level physics instruction and reported greater learning in less time than in a comparison active-learning class, together with favorable student perceptions. The intervention incorporated established instructional practices rather than allowing unrestricted answer generation. Its findings counter the notion that assistance necessarily suppresses learning. AI can deliver effective scaffolding, especially when the learning process—not merely the final output—is what the system is configured to support.

The wider evidence base is mixed rather than uniformly negative. Chen and Cheung's (2025) meta-analysis synthesized 57 studies of university learning and reported positive average effects on several outcomes, including academic achievement and higher-order thinking, but no statistically significant average effect on metacognition. Such pooled results are informative about broad tendencies, yet their

educational interpretation depends on study quality, outcome definitions, duration, and whether assessments required independent performance. A positive average effect on measured reasoning tasks does not prove a long-term increase in intellectual autonomy; equally, a null metacognitive estimate does not prove deterioration.

More recent longitudinal and synthesis work reinforces the importance of context. Melanou et al. (2026), examining three instructional conditions across a semester with 87 students, found gains in knowledge across conditions but no additional overall benefit from AI integration; reflective use was associated with critical thinking. A scoping review by Lin et al. (2026) mapped 394 empirical higher-education studies and noted that much of the field still relies on short-term, readily measured outcomes, while self-regulated learning and equity receive comparatively less attention. Since a scoping review maps research rather than establishing causal efficacy, it should not be interpreted as a verdict on whether AI enhances or erodes thinking.

Evidence from outside classrooms can clarify mechanisms without proving student outcomes. Lee et al. (2025) surveyed 319 knowledge workers and found that greater confidence in GenAI was associated with less self-reported critical-thinking effort, whereas greater confidence in one's own abilities was associated with more. Respondents described a shift from information generation toward verification and integration. Because the evidence is self-reported and drawn from work settings, it cannot establish that students' critical-thinking abilities decline. It nevertheless highlights a potentially important transition: cognition may change its location and character rather than simply increase or disappear.

Taken together, these findings justify a conditional proposition. Unstructured AI assistance can reduce the amount of independently performed reasoning in an educational task, and in some settings this can impair later unaided performance. Well-designed AI assistance can support learning, sometimes substantially. Whether repeated patterns produce enduring changes in initiative, intellectual confidence, or domain-general critical-thinking ability remains insufficiently established. Claims about permanent decline should therefore be treated as research hypotheses rather than settled facts.

V. THREE PATHWAYS THROUGH WHICH INDEPENDENT THINKING MAY WEAKEN

5.1 Premature closure: receiving a question's answer before constructing the question

The first pathway is premature closure of inquiry. Many educational tasks are valuable precisely because students must determine what is being asked, decide what information matters, and consider competing approaches. An AI response can collapse these stages into a ready-made interpretation before the student has experienced the uncertainty that motivates inquiry. A learner assigned to discuss whether a policy is just, for example, may ask a chatbot for a balanced essay and receive familiar arguments on both

sides. What is lost is not only writing practice; it is the prior work of deciding which competing definitions of justice are relevant and what evidence would distinguish them.

This risk is especially important in disciplines where framing is a major component of expertise. Historical interpretation requires selecting and contextualizing sources; scientific research requires generating testable questions; literary criticism requires attending to ambiguity; education research requires distinguishing normative claims from causal explanations. An AI answer can be useful after students have made an initial interpretive commitment. Delivered too early, it may narrow the space of possibilities they would otherwise have explored.

5.2 Epistemic passivity: mistaking plausibility for justification

The second pathway is epistemic passivity. Generative systems are optimized to produce plausible responses, not to guarantee the truth of every statement or the adequacy of every inference. In academic work, however, a claim becomes acceptable through evidence, valid reasoning, and accountability to disciplinary standards. Students who treat AI output as a convenient authority may skip the intermediate question: why should this conclusion be believed?

A sophisticated form of passivity can hide behind extensive editing. Students may reorganize paragraphs, change examples, or improve style while retaining an unsupported premise supplied by AI. Observable activity is therefore not sufficient proof of critical engagement. Evaluative reasoning requires confronting substantive alternatives, searching for disconfirming evidence, and identifying conditions under which a conclusion would change. It is particularly demanding when AI output matches the learner's prior beliefs or appears in authoritative academic language.

This risk also changes what teachers must teach. Traditional information literacy frequently emphasizes evaluating sources and distinguishing scholarly from non-scholarly material. GenAI requires an additional level of scrutiny because individual sentences can be fluent yet untraceable to a verifiable source. Independent thinkers must ask whether a reference exists, whether quoted claims accurately represent it, whether opposing evidence has been excluded, and whether a persuasive synthesis conceals conceptual inconsistency.

5.3 The erosion of self-monitoring through repeated rescue

The third pathway is the repeated external correction of difficulties that students could learn to manage. When an AI tutor immediately detects every error and supplies the next action, students may become less practised at recognizing when a strategy has failed. Over time, the learner's internal question—"What am I misunderstanding?"—can be replaced by the operational question—"What should I ask the chatbot next?" The latter can be efficient, but it is not a substitute for the former.

Evidence from a year-long behavioral-sequence study by Dai et al. (2026) is suggestive: higher-performing graduate-student groups maintained more critical patterns of AI engagement, whereas lower-performing groups showed patterns consistent with greater cognitive outsourcing. Because these patterns were observed at the group level and do not establish causation, they cannot prove deterioration produced by AI. They do, however, indicate why averages can conceal divergent developmental trajectories. Some students learn to interrogate the technology; others may increasingly delegate the work of interrogation.

These pathways can reinforce one another. Premature answers reduce opportunities for initial hypothesis formation; persuasive outputs lower the perceived need for evaluation; continuous correction reduces practice in monitoring. Together, they may gradually alter classroom norms about what counts as learning. The critical issue is not the number of prompts students send, but whether interaction with AI strengthens their ability to act without it.

VI. WHY GENAI CAN ALSO STRENGTHEN ACTIVE AND INDEPENDENT THINKING

A credible argument must address the opposite possibility: AI can enlarge intellectual agency when it makes previously inaccessible forms of inquiry possible. Students often experience barriers that are not themselves the intended learning target. Limited vocabulary, unfamiliar terminology, language anxiety, inaccessible materials, or delayed teacher feedback can prevent them from engaging in demanding conceptual work. Responsive explanations and adaptive hints can reduce these barriers. When assistance helps learners enter a task rather than complete it for them, cognitive offloading is potentially enabling.

A second opportunity is dialogic challenge. A conventional textbook cannot readily question a student's reasoning or generate counterexamples tailored to an initial claim. A carefully constrained AI interlocutor can ask, “What evidence would disprove this interpretation?” or “How does this conclusion change if the underlying assumption is false?” Such exchanges can sustain constructive engagement when the student must answer in their own terms. In this use, the technology serves as a source of productive resistance rather than a supplier of finished judgments.

Third, AI can support formative assessment. Teachers face limited time, especially in large classes, and students may receive feedback only after the opportunity for revision has passed. AI-supported feedback can identify places where explanations are incomplete or provide practice questions targeted to misunderstandings. For these benefits to count as learning rather than mere correction, students must interpret the feedback, explain what they changed, and later demonstrate that they can apply the principle independently.

Finally, responsible AI use can itself become an object of critical education. Students can learn how probabilistic language systems generate text, why answers vary with prompts, what data biases can shape outputs, and why apparent confidence is not evidence. UNESCO's AI competency framework emphasizes a

human-centered mindset alongside ethical and technical understanding (Miao et al., 2024). These capabilities are increasingly relevant to informed participation in social and professional life. An educational response focused only on prohibition would leave students less prepared to make responsible decisions in environments where AI is widespread.

The educational objective is therefore not maximal student effort or minimal AI use. It is a defensible division of labor: AI should reduce barriers to learning while preserving, and sometimes intensifying, the forms of reasoning students need to master.

VII. PROTECTED COGNITIVE AGENCY: A CONDITIONAL PRINCIPLE FOR AI INTEGRATION

The argument so far supports a guiding principle: **protected cognitive agency**. This term is used here to describe an instructional commitment, not a newly validated psychological scale or a theory claiming unique explanatory priority. The principle holds that learners should retain meaningful control over four responsibilities: defining the intellectual problem, developing an initial account, evaluating competing reasons, and demonstrating that understanding can be reconstructed without the system. Assistance may be extensive, but these responsibilities should not be routinely surrendered when they constitute the educational target.

The principle has three dimensions. The first is **temporal**: when is the system introduced? Assistance before any initial attempt can preempt the student's generation of hypotheses; assistance after a genuine attempt can serve as informative feedback. The second is **functional**: which operation is assigned to the system? Automating citation formatting differs from automating source evaluation; translating instructions differs from choosing a research question. The third is **accountability-based**: can the student explain, test, and transfer the resulting knowledge without simply repeating the system's wording? A fourth, **developmental**, dimension concerns how these choices change as a learner gains expertise. Novices may require more explanation and structured support, whereas advanced students should progressively assume responsibility for evaluating, adapting, and challenging it.

From these dimensions, a practical sequence can be derived: **attempt, consult, contest, reconstruct**. In the attempt phase, students first generate a question, provisional claim, worked step, or explanation without AI assistance. This initial work establishes an independent intellectual position and reveals uncertainty. In the consult phase, students use AI for targeted feedback, hints, comparison, or alternative perspectives rather than simply requesting a completed answer. In the contest phase, they actively test the output against evidence, identify errors or assumptions, and explain why they accept or reject suggestions. In the reconstruct phase, they produce an account or solve a related problem without access to the earlier AI response.

The sequence is a pedagogical heuristic, not a claim that every subject or learner should follow an identical rigid procedure. In an accessibility context, initial AI support may be necessary before a student can make any meaningful attempt. In advanced research, consultation and contestation may recur repeatedly rather than proceed linearly. Nevertheless, the four phases specify what a teacher can observe and what an assessment can require. They shift attention from whether AI was used to how the student exercised judgment around its use.

VIII. IMPLICATIONS FOR TEACHING: DESIGNING TASKS THAT REQUIRE THOUGHT

The first implication is to design learning activities around intellectual decisions rather than only deliverable products. Asking students to submit an essay on a familiar topic is increasingly insufficient evidence of independent reasoning. Asking them to formulate competing theses, justify the selection of evidence, identify weaknesses in a model explanation, or explain how a counterexample changes their position makes the process more visible. Teachers need not prohibit AI to achieve this; they can specify which decisions must be made by students and which supports are permissible.

Second, AI interaction should be scaffolded according to the learning objective. When the goal is mathematical reasoning, a tutoring interface can reveal a hint or ask a diagnostic question before providing a complete solution. When the goal is source criticism, the system may be asked to produce competing claims that students must substantiate using original documents. When the goal is writing style, grammar assistance may be appropriate, but students should still establish the claim and justify its evidential basis. The key design test is whether the learner must perform the cognitive operation that the lesson claims to develop.

Third, teachers should normalize constructive uncertainty. Students may turn to AI prematurely because they interpret hesitation as evidence that they are incapable of succeeding. Short periods of unaided brainstorming, prediction, or explanation can make uncertainty a routine part of learning. These phases need not be punitive or excessively long. A three-minute independent attempt followed by a structured comparison with AI feedback may be more educationally useful than either unlimited struggle or immediate disclosure of the answer.

Fourth, classrooms should explicitly teach verification. A useful task is to give students an AI-generated explanation containing a mixture of accurate statements, incomplete reasoning, and a plausible but unsupported claim. Students must identify what can be accepted, what requires evidence, and what must be corrected. However, such exercises should not reduce verification to a game of spotting intentionally planted mistakes. In authentic inquiry, the more demanding question is often how to evaluate an answer that is mostly correct but conceptually incomplete. Students need criteria drawn from the relevant discipline, not only generic skepticism.

Fifth, teachers should require reflection that goes beyond reporting which tool was used. Process notes can ask students to identify their initial position, describe a significant disagreement with AI, explain the evidence that resolved it, and specify what they would do differently in the next task. A disclosure statement establishes transparency; a reasoned reflection can reveal learning. The two functions should not be confused, and neither should become a bureaucratic substitute for actual reasoning.

Sixth, peer discussion remains valuable. AI conversations are personalized, responsive, and private, but education also depends on encountering real disagreement and negotiating meaning with other people. Small-group defense of alternative solutions can reveal assumptions that individuals overlook. Students may bring AI-generated counterarguments into discussion, but responsibility for evaluating them should remain social and human. UNESCO's guidance for GenAI in education emphasizes protecting human agency and maintaining opportunities for independent reasoning and interpersonal learning (Miao & Holmes, 2023).

Seventh, educators need realistic implementation rules. A blanket demand to verify every sentence would impose an unsustainable burden, and absolute bans may be unenforceable or inequitable. A more workable policy distinguishes low-stakes assistance, substantive intellectual contribution, and high-stakes evaluation. Students could use AI for accessible explanations or language support while disclosing substantial contributions to argumentation or analysis. Teachers, in turn, should make clear why certain stages require independent work, connecting restrictions to learning goals rather than treating every AI interaction as misconduct.

IX. RETHINKING ASSESSMENT: FROM POLISHED PRODUCTS TO DEMONSTRABLE UNDERSTANDING

Traditional assessment can become misleading when the quality of an artifact no longer reliably indicates the competence of the person submitting it. This does not make essays, projects, or take-home assignments obsolete. It requires greater clarity about what they measure. A well-written essay may demonstrate a successful human–AI production process, but it cannot by itself establish the student's unaided analytical proficiency. If institutions continue to infer the latter from the former, they may reward students for work they have not learned to perform.

A defensible assessment system can combine supported and unsupported performance. Students may develop a project with AI, while also completing a short independent explanation, a transfer problem, or an oral defense. The objective is not to catch students using technology; it is to establish whether the learning goals have been achieved. For some disciplines, independent recall and procedural fluency are essential. For others, the central competence is evaluating evidence, making ethical judgments, or applying concepts

under uncertainty. The assessment method should follow the construct rather than a uniform rule about technology.

Process-sensitive evaluation can also help. Version histories, brief reasoning journals, annotated decision points, and selective interviews may reveal whether students developed an argument or merely accepted one. Yet extensive process surveillance raises privacy concerns and can transform education into performative compliance. The purpose should be to sample meaningful cognitive decisions, not to collect exhaustive records of every keystroke or prompt. Students should understand what evidence is being requested and why.

A useful assessment distinction is between **performance with AI**, **immediate performance without AI**, and **delayed transfer without AI**. The first describes assisted productivity; the second indicates current accessible competence; the third provides stronger evidence of retained and transferable learning. All three are educationally relevant, but they answer different questions. A course claiming to develop independent reasoning should not rely exclusively on the first measure. The experimental contrast used by Bastani et al. (2025) illustrates why the removal of assistance can reveal outcomes that remain invisible during supported practice.

Fairness also matters. Students differ in access to paid tools, stable internet connections, prior AI literacy, language proficiency, and support from teachers. Assessment should not assume identical technological opportunities or punish students whose accommodations legitimately involve assistive software. Institutions should articulate assessment conditions in advance and provide accessible alternatives. The objective is a valid interpretation of learning, not an idealized image of unsupported work that disregards ordinary educational resources.

X. EQUITY, INSTITUTIONAL INCENTIVES, AND THE HIDDEN CURRICULUM OF EFFICIENCY

AI integration does not occur in a neutral environment. Students face grade competition, scholarship requirements, overloaded schedules, and assignments that may reward conformity more visibly than curiosity. If institutions assess only finished products, students have rational incentives to use tools that improve those products quickly. It is therefore too simple to blame a perceived decline in independent thinking on individual students' lack of discipline. Educational structures can make cognitive outsourcing the efficient response to the incentives they create.

The benefits and risks are also distributed unevenly. Students with stronger prior knowledge are often better positioned to notice weak reasoning, formulate informative prompts, and use AI as a critic. Those with limited foundational understanding may find plausible but flawed answers more difficult to challenge. This creates the possibility that the same technology could strengthen the agency of already-advantaged

learners while increasing dependence among others. The possibility deserves investigation rather than assumption: Melanou et al. (2026), for example, found no evidence of an increased performance divide in their relatively small semester-long study. Unequal outcomes should be treated as a conditional risk, not an inevitable result.

At the same time, restricting assistance can disadvantage learners who use AI to overcome language barriers, disability-related constraints, or limited access to human tutoring. An equitable policy must therefore separate the intellectual work that defines the learning objective from barriers that are irrelevant to that objective. If students are learning causal inference, grammar correction may improve fairness; if they are learning academic English composition, writing assistance may need clearer limits. The relevant distinction is pedagogical, not merely technological.

Institutions should also reconsider productivity-centered language. If students repeatedly hear that successful learning means producing more pages, answering faster, and eliminating uncertainty, they may internalize the idea that thinking is an avoidable cost. A more defensible educational culture values the quality of questions, the revision of beliefs, and the capacity to provide reasons. GenAI makes that cultural decision urgent because it can produce an appearance of mastery at unprecedented speed.

XI. A RESEARCH AGENDA FOR TESTING CLAIMS OF COGNITIVE DECLINE

The most important unresolved question is longitudinal. Does repeated use of GenAI change students' independently demonstrated reasoning capacities or their willingness to exercise them across months and years? A short experiment can show effects on a particular task, but enduring developmental claims require stronger evidence. Researchers should distinguish temporary performance dependence, loss of practice in a narrow skill, generalized transfer deficits, and changes in motivational habits. These are related but not interchangeable outcomes.

A first priority is to manipulate the **division of cognitive labor**, not merely AI access. Studies could randomly assign learners to unaided work, unrestricted answer generation, hint-based tutoring, and an attempt–consult–contest–reconstruct procedure. The interfaces should be matched as far as possible for time, content availability, and usability. Such designs would test whether preserving initial generation and subsequent reconstruction changes independent learning beyond the benefits of receiving feedback. Pre-registration and fidelity checks would help prevent selective interpretation of complex outcomes.

A second priority is multimethod measurement. Independent-thinking dispositions can be assessed through validated questionnaires, but surveys alone cannot demonstrate reasoning quality. Behavioral tasks should measure spontaneous question generation, argument evaluation, error detection, explanation revision, and strategic persistence. Process traces can show when students seek help and whether they examine responses critically. Interviews may clarify students' intentions. Crucially, measurement models should

distinguish cognitive initiative, reasoning ability, metacognitive monitoring, and intellectual confidence rather than treating them as interchangeable indicators of a single poorly defined outcome.

A third priority is to evaluate delayed and novel performance. Immediate post-tests may reflect familiarity with examples seen during AI-supported practice. Researchers should therefore include delayed tests and tasks requiring transfer to unfamiliar cases. When possible, they should assess both performance with and without AI. The difference between these conditions provides a direct indicator of reliance on assistance, although even that difference must be interpreted relative to task difficulty and the learner's prior knowledge.

A fourth priority is to examine developmental and disciplinary moderators. Younger learners, novices, students learning in a second language, and advanced researchers face different cognitive demands. In mathematics, worked steps may be central to learning; in historical research, source evaluation and contextual interpretation may be more important; in creative writing, developing an original voice may be the target. A single universal rule for acceptable offloading would ignore these differences. Experimental designs should explicitly specify which operations are peripheral and which constitute the learning goal.

Finally, research should include implementation failures and unequal effects. Carefully engineered AI tutors can perform differently from freely available systems that students use without instructions. Studies conducted with enthusiastic instructors may not generalize to crowded classrooms. Tool updates can also alter behavior during a trial, making transparent reporting essential. Rather than asking for a single average answer to whether AI weakens thinking, the field should specify which learners, tools, pedagogies, and assessment systems produce which developmental trajectories. That is a more demanding research agenda, but it is also more educationally useful.

XII. CONCLUSION: THE RIGHT TO THINK, AND THE RESPONSIBILITY TO TEACH THINKING

The concern that generative AI may weaken students' independent thinking is legitimate, but its strongest formulation is not that using AI inevitably makes students less intelligent or less capable. The more defensible concern is that some forms of AI use bypass repeated opportunities to formulate questions, construct explanations, evaluate evidence, and monitor error. When educational systems reward finished answers more than the reasoning needed to produce them, powerful generative tools can intensify an existing weakness in pedagogy.

The empirical record does not support a general claim of irreversible cognitive decline. It shows that unrestricted assistance can undermine later independent performance in some settings, while well-designed tutoring can improve learning in others. Theoretical accounts of cognitive offloading, self-regulation, and constructive engagement explain why these apparently contradictory outcomes can coexist. Their common

lesson is that cognition is not protected merely by reducing screen time or maximizing visible effort. It is protected when the learner continues to perform the intellectual operations necessary for growth.

The proposed principle of protected cognitive agency gives this lesson a practical form. Students should be supported in attempting problems, consulting resources, contesting claims, and reconstructing understanding. Teachers should evaluate independent competence alongside AI-assisted productivity, and institutions should align incentives with genuine learning rather than polished output alone. The decisive educational question is therefore not whether students are permitted to use artificial intelligence, but whether they are being taught to remain the authors, critics, and responsible owners of their own judgments. The future of independent thinking will depend less on the presence of AI in classrooms than on the educational choices made around it.

REFERENCES

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

Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis. *Educational Research Review, 49*, 100737. https://doi.org/10.1016/j.edurev.2025.100737

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Dai, Y., Chen, F.-Y., Xiao, J.-Y., Fan, A. C. W., Guo, J.-W., & Zhang, M. (2026). Cognitive outsourcing in generative AI-supported learning: A behavioural-sequence analysis of metacognitive differentiation. *SAGE Open*. https://doi.org/10.1177/21582440261497399

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, 17458. https://doi.org/10.1038/s41598-025-97652-6

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Article 1121). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Lin, M. P.-C., Chang, D. H., Winne, P. H., Bowen, G. M., & Ryoo, J. (2026). Mapping generative AI in higher education teaching and learning: A scoping review. *International Journal of Educational Technology in Higher Education, 23*, 55. https://doi.org/10.1186/s41239-026-00630-9

Melanou, C., Beege, M., & Kimmig, M. (2026). Generative AI and learning dynamics in higher education: A longitudinal empirical study. *Journal of Computer Assisted Learning, 42*(5), e70322. https://doi.org/10.1002/jcal.70322

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO. https://www.unesco.org/en/articles/ai-competency-framework-students

Nakamura, D. (2026). Augmentation or erosion? Regulating cognitive offloading in generative AI-supported STEM learning. *Educational Psychology Review, 38*, 124. https://doi.org/10.1007/s10648-026-10224-6

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

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, 108386. https://doi.org/10.1016/j.chb.2024.108386

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2