# A COGNITIVE-PRESERVATION FRAMEWORK FOR YOUNG LEARNERS IN THE AGE OF GENERATIVE AI

Ronald Jean Degen

ANEGEPE (Academia de Estudos em Empreendedorismo e Gestão de Pequenas Empresas) São Paulo, Brazil

## *ABSTRACT*

*Generative artificial intelligence (AI) can expand access to explanation, feedback, and personalized support, but it can also perform cognitive operations through which learning is built. This conceptual review examines how young learners can use generative AI without outsourcing retrieval, reasoning, argumentation, verification, and transfer. A large field experiment in high-school mathematics found that unrestricted GPT-4 assistance improved practice performance but reduced later independent performance, whereas pedagogical guardrails largely mitigated this effect. Other recent evidence is mainly correlational or preliminary and does not justify broad causal claims. Drawing on cognitive offloading, cognitive load theory, retrieval practice, generative learning, metacognition, and emerging AI studies, the article proposes a three-phase Cognitive-Preservation Framework: think before AI, think with AI, and think again without AI. The framework positions AI as scaffold, tutor, critic, and intellectual adversary while preserving productive mental effort.*

## *KEYWORDS*

*Artificial intelligence, Cognitive offloading, Critical thinking, Generative AI, Learning*

## 1. INTRODUCTION

Generative artificial intelligence has moved rapidly from a specialist technology to an ordinary intellectual tool. Students can now ask a conversational system to explain a concept, summarize a chapter, generate examples, translate a passage, propose an argument, write computer code, solve an equation, create a study plan, or draft an essay within seconds. In educational settings, this capacity has obvious benefits. It can reduce barriers to information, provide immediate feedback, adapt explanations to a learner's prior knowledge, and offer additional practice at a scale that is difficult to achieve through human instruction alone. At the same time, generative AI differs from earlier educational technologies in a crucial respect: it does not merely store or retrieve information. It can perform a substantial portion of the reasoning, synthesis, planning, and expression that educators traditionally ask students to perform in order to learn.

The educational question is therefore no longer whether students will use AI. The more consequential question is how they should use it. A prohibition-only approach is unlikely to be sustainable, yet unrestricted use risks transforming learning tasks into answer-acquisition tasks. When a student delegates the central intellectual operation of an assignment to a model, the final product may improve even while the learner's independent capability changes little. The distinction between visible performance during assisted practice and durable learning after assistance is removed is central to the science of learning [1]. Generative AI makes that distinction especially important because it can raise the quality and speed of a student's output so dramatically that assisted performance can be mistaken for competence.

The present article develops a cognitive-preservation approach to generative AI in education. Its central proposition is simple: AI should reduce cognitively peripheral burdens while preserving the productive mental effort that constructs knowledge. This proposition builds on research on cognitive offloading [2], cognitive load theory [3], [4], retrieval practice [5], [6], generative learning [7], and the difference between short-term performance and long-term learning [1]. It is also informed by emerging studies that document task-specific benefits of guarded AI assistance, dependence-related associations, and risks of poorly guarded use [8]-[12].

The argument is not that cognitive effort is intrinsically virtuous or that students should reject technological assistance. Human cognition has long used external supports to reduce memory and processing demands [2], [13]. Rather, the pedagogical problem is one of selective delegation. The operations that may be safely automated depend on the capability being developed. If the learning goal is argumentation, delegating the argument reduces direct practice in argument construction. If the learning goal is programming logic, requesting complete code before attempting the problem can bypass the target process. If the learning goal is understanding a scientific concept, asking for a simpler explanation may be helpful; asking the system to complete every inference may not be. The educational task is therefore to distinguish assistance that scaffolds cognition from assistance that substitutes for the cognitive operation being trained.

## 2. PURPOSE AND CONCEPTUAL APPROACH

This article is a conceptual review rather than a systematic review or original empirical study. It synthesizes established findings from cognitive and educational psychology with recent research on generative AI in education. The purpose is to construct a theoretically coherent framework that can guide student behavior and instructional design while the empirical literature on long-term cognitive consequences remains relatively young. To reduce evidentiary overreach, claims are calibrated to study design: experimental findings are used for task-specific causal inferences; cross-sectional studies are described as associations rather than causes; opinion and review articles are used for conceptual synthesis; and preprint evidence is explicitly identified as preliminary. Particular attention is given to mechanisms already well established independently of generative AI—working-memory limits, cognitive offloading, retrieval, generative learning, metacognitive monitoring, and the learning-performance distinction—and to ways generative AI may alter how often learners engage those mechanisms.

Four questions organize the analysis:

**First**, why can AI improve immediate performance without necessarily improving learning?

**Second**, which forms of cognitive offloading are educationally useful and which are likely to undermine skill formation?

**Third**, how can AI reduce unnecessary cognitive load without removing the mental processing required to construct schemas and transferable knowledge?

**Fourth**, what practical routines can preserve independent reasoning while still exploiting the explanatory, feedback, and dialogic advantages of generative AI?

## 3. THE COGNITIVE PARADOX OF GENERATIVE AI

Jose et al. [9], in an Opinion article that synthesizes emerging literature rather than reporting an original experiment, describe a cognitive paradox in which AI can function as both an amplifier and an inhibitor of human cognition. The paradox is not a contradiction in the technology itself; it reflects different patterns of use. The authors argue that AI may support cognition when it provides personalization, feedback, questions, and alternative perspectives, but may erode

cognitive engagement when it is used as a substitute for reasoning. In the present article, this source is therefore used as a conceptual framing device, not as causal evidence.

This duality is consistent with the broader distinction between performance and learning. Performance is what can be observed during practice; learning is a relatively durable change in knowledge or skill that can support later retention and transfer. These measures can diverge. Soderstrom and Bjork [1] review evidence showing that conditions that make practice look fluent can sometimes produce weaker long-term learning, while conditions that introduce manageable difficulty can impair immediate performance yet strengthen later retention. Generative AI intensifies the risk of confusing these outcomes because the student-plus-AI system can produce a polished product even when the student alone could not reproduce its reasoning.

A particularly clear demonstration comes from Bastani et al. [8], who conducted a field experiment with nearly one thousand high-school students learning mathematics. During practice, grades improved by 48% with the unrestricted GPT Base interface and by 127% with the guarded GPT Tutor relative to the no-AI control. When AI access was subsequently removed, students who had used GPT Base scored 17% lower than students who had never had access to AI. The negative learning effect was largely mitigated in GPT Tutor, which was designed to constrain direct answer-giving and provide teacher-designed hints. The result supports a task-specific principle: performance achieved with a tool present is not sufficient evidence that the underlying skill has been acquired.

This principle should change how students, teachers, and institutions interpret productivity gains. Faster homework completion, more sophisticated prose, cleaner code, or higher-quality problem sets may reflect genuine learning, effective scaffolding, simple substitution, or some mixture of all three. The educational objective is not merely to maximize output quality under assisted conditions. It is to develop capabilities that remain available when assistance is absent, unreliable, inappropriate, or insufficient.

## 4. COGNITIVE OFFLOADING: WHEN A TOOL BECOMES A SUBSTITUTE

Cognitive offloading refers to the use of external actions or resources to reduce the information-processing demands of a task [2]. Offloading is a normal feature of intelligent behavior. People write notes so that they do not need to remember every appointment, use calculators to avoid repetitive arithmetic, and rely on maps to reduce spatial-memory demands. Digital technologies expanded this process: the expectation that information will remain externally available can shift memory toward remembering where to find information rather than remembering the information itself [13].

Generative AI broadens offloading from memory and calculation to higher-order processes. A user can now externalize idea generation, planning, synthesis, comparison, argument construction, editing, coding, and problem solving. This does not make offloading inherently harmful. In many professional settings, delegating routine operations can free cognitive resources for more consequential decisions. In education, however, the value of offloading depends on whether the externalized operation is peripheral to the learning goal or is itself the capability being trained.

Consider two students studying the causes of a historical event. The first writes what she already knows, organizes tentative causes, identifies uncertainties, and then asks AI to challenge her causal model. The second asks AI to provide the causes immediately and rewrites the response. Both students receive useful information, but the first is required to retrieve prior knowledge, generate a representation, detect gaps, compare explanations, and revise a model. Those

operations are not incidental; they are part of learning. The second student may produce a better-looking answer while practicing fewer of the target cognitive processes.

The same principle applies across disciplines. A mathematics learner can ask for a hint after attempting a problem rather than requesting the complete solution at the outset. A programming learner can ask the model to diagnose why a specific function fails after writing code, rather than requesting an entire application. A writer can generate a first argument and ask the model for counterarguments, rather than asking the model to invent the thesis and structure. In each case, the technology assists after the learner has engaged the mental operation that instruction is intended to develop.

The practical rule is therefore not "do not offload." It is "do not offload the target capability." A learner must first identify the capability being developed—recall, explanation, reasoning, writing, modeling, coding, evaluation, or decision-making—and then ask whether the AI is supporting that capability or replacing its exercise. This distinction provides a more precise criterion than simply classifying AI use as permitted or prohibited.

## 5. COGNITIVE LOAD THEORY: REDUCE THE WRONG LOAD, PRESERVE THE RIGHT EFFORT

Cognitive load theory begins from a basic feature of human cognition: working memory is limited when processing novel information, whereas organized knowledge stored in long-term memory can support complex performance with far greater efficiency [3], [4]. Instruction should therefore avoid unnecessary demands that consume working-memory resources without contributing to learning. Generative AI can be highly useful in this respect. It can clarify confusing directions, restate a difficult explanation, provide an example, translate unfamiliar terminology, organize data, or adapt the level of language to a learner's current knowledge.

A common misinterpretation, however, is to infer that good educational technology should minimize mental effort in general. Learning does not occur simply because information is available in an easy-to-consume form. Generative learning requires learners to select relevant information, organize it into a coherent representation, and integrate it with prior knowledge [7]. Retrieval is a distinct learning process that can itself strengthen later retention and flexible access to knowledge [5], [6]. Fiorella and Mayer [7] identify strategies such as summarizing, mapping, self-testing, self-explaining, and teaching as ways of inducing generative processing.

The crucial distinction is therefore between reducing extraneous processing and eliminating generative processing. AI can lower barriers that are irrelevant to the target objective, but it should not automatically remove the reasoning burden that constitutes the objective. If a student is learning physics, the model may help interpret notation or provide an analogous example, but the student still needs to reason about physical relationships. If the student is learning academic writing, the model may identify redundancy or unclear transitions, but the student should still formulate the thesis, select evidence, and organize the logic. If the student is learning statistics, the model may explain a software error, but the student should still decide which analysis answers the research question and why.

This reframing changes the question students should ask. Instead of asking, "How can AI make this task as easy as possible?" the student asks, "Which part of this difficulty is irrelevant to the capability I am trying to build, and which part is the exercise that builds it?" Productive educational use of AI depends on preserving the second category even when the first can be automated.

## 6. METACOGNITION, RETRIEVAL, AND THE ILLUSION OF COMPETENCE

Generative AI can also create a metacognitive problem. The learning-performance literature shows that smooth or successful performance during practice can be an unreliable indicator of durable learning [1]. In an AI-assisted context, a learner who reads a clear generated solution may feel that the concept is obvious, yet recognition while an explanation is visible differs from the ability to reconstruct the explanation independently. The risk is therefore not that fluent explanations are inherently harmful, but that fluency may be mistaken for evidence that learning has already occurred.

Retrieval practice provides a direct corrective. Actively attempting to recall information strengthens long-term retention and can improve later access to knowledge [6]. Karpicke and Blunt [5] further showed that retrieval practice can support meaningful learning and conceptual understanding, not merely rote recall. Dunlosky et al. [14], reviewing a range of study techniques, identified practice testing and distributed practice among the most broadly useful strategies. These findings have immediate implications for AI-assisted study: after receiving an explanation, the learner should remove the explanation and attempt to reconstruct it from memory.

A simple test can separate access from ownership: “Can I explain this without looking?” If the answer is no, the learner has gained information access but not yet demonstrated independent knowledge. Stronger tests ask the learner to solve a new problem, generate an example, explain the concept to another person, identify a boundary condition, or apply the idea in a different context. These activities force retrieval and transfer, reducing the risk that the fluency of the AI output becomes an illusion of personal competence.

Metacognitive monitoring also requires verification. AI dialogue systems can generate inaccurate or misleading information, a risk emphasized in the literature on over-reliance [12]. More directly, Walters and Wilder [15] evaluated 636 bibliographic citations produced by ChatGPT-3.5 and GPT-4 and documented both fabricated references and substantive errors in real citations. The educational implication is not that every AI output is unreliable, but that fluent prose and plausible-looking references cannot substitute for checking claims against primary or otherwise authoritative sources.

## 7. EMERGING EVIDENCE ON AI DEPENDENCE AND COGNITIVE ENGAGEMENT

The empirical literature on the long-term cognitive effects of generative AI is still developing, and strong causal claims should be avoided. Zhai et al. [12] systematically reviewed 14 studies published from 2017 to 2023 on over-reliance on AI dialogue systems in educational and research contexts. The review synthesized literature linking over-reliance and uncritical acceptance of AI output with concerns about decision-making, critical thinking, and analytical reasoning. Because the underlying corpus was small, methodologically diverse, and not designed to establish long-term causal cognitive decline, the review is best interpreted as evidence of a risk pattern requiring further study rather than proof that AI use itself diminishes cognition.

Tian and Zhang [10] collected cross-sectional survey data from 580 Chinese university students. Greater AI dependence was associated with lower critical-thinking scores; cognitive fatigue statistically mediated part of this association, and information literacy moderated parts of the model. These findings are consistent with a possible dependence-related pathway, but cross-sectional mediation cannot establish temporal or causal mechanism. The study therefore supports

an association between AI dependence, fatigue, and critical-thinking measures, not the stronger claim that AI dependence causes a decline in critical thinking.

More recent work by Tian and Zhang [11] further illustrates the double-edged character of AI engagement. In a cross-sectional study of 698 Chinese undergraduates, perceived AI intelligence was positively associated with critical thinking, both directly and through focused immersion. A negative sequential association also appeared: higher perceived AI intelligence was associated with greater focused immersion, focused immersion with greater AI dependency, and dependency with lower critical thinking. Because the data are cross-sectional, these paths should not be read as a temporal chain in which immersion causes dependency or dependency causes cognitive decline. Rather, the study identifies potentially beneficial attentional and detrimental reliance pathways that require longitudinal and experimental testing.

Kosmyna et al. [16] examined essay writing under LLM-assisted, search-engine, and unaided conditions using behavioral measures and electroencephalography. The preprint reports 54 participants in Sessions 1-3 and 18 in a fourth crossover session. The LLM group showed the weakest overall EEG connectivity, the lowest self-reported ownership of essays, and greater difficulty accurately quoting its own writing than the unaided group. These findings are preliminary and should not be equated with brain damage, reduced intelligence, or a demonstrated long-term loss of cognitive capacity. A methodological comment by Stanković et al. [17] raises concerns about sample size, reproducibility, EEG analysis, reporting consistency, and procedural transparency. Accordingly, the study is treated here as hypothesis-generating evidence about cognitive engagement during a specific writing task, not as definitive evidence of cognitive deterioration.

Taken together, the evidentiary hierarchy matters. Bastani et al. [8] provide direct field-experimental evidence that unrestricted answer-oriented AI can improve assisted performance yet impair subsequent unaided performance in high-school mathematics, and that pedagogical guardrails can largely mitigate that effect. Tian and Zhang [10], [11] provide cross-sectional evidence that AI dependency is associated with lower critical-thinking measures, but they do not establish causality. Kosmyna et al. [16] provide preliminary neurobehavioral observations from a preprint whose methods have been challenged [17]. A restrained conclusion is therefore warranted: interaction design and patterns of dependence can matter for learning, but broad claims that generative AI generally weakens cognition exceed the present evidence.

## 8. A COGNITIVE-PRESERVATION FRAMEWORK FOR AI-ASSISTED LEARNING

The principles above can be operationalized through a three-phase routine: think before AI, think with AI, and think again without AI. The purpose is to preserve the cognitive operations that build competence while still exploiting the speed, feedback, and flexibility of generative systems. The framework is intentionally simple so that it can be used by adolescents, university students, teachers, and families across disciplines.

### 8.1. Think Before AI: Generate An Initial Cognitive Representation

Before opening a chatbot, the learner should perform a small amount of independent cognitive work. The goal is not to solve the entire task without assistance. The goal is to activate relevant prior knowledge, reveal uncertainty, and create an initial representation that can later be compared with external feedback. Depending on the task, the learner can write what is already known, attempt the first step of a problem, sketch an outline, make a prediction, formulate a

hypothesis, identify a confusing point, or write an initial argument. This phase is consistent with generative learning because it requires the learner to select, organize, and connect information rather than begin from a finished external answer [7].

This phase prevents AI from becoming the first source of thought. For example, instead of asking, “Write an essay about climate policy,” a student first writes a provisional thesis and two supporting reasons, then asks the model to identify weaknesses and counterarguments. Instead of asking for a complete solution to an algebra problem, the student attempts the setup and asks whether the chosen method is appropriate. Instead of asking for a program, the student writes pseudocode and asks the AI to test the logic. The learner’s initial attempt creates a cognitive anchor and makes subsequent feedback diagnostically useful.

### 8.2. Think With AI: Use The System As Tutor, Critic, And Socratic Partner

During AI interaction, the goal should be to convert the system from an answer machine into a thinking partner. Prompts designed to preserve learner-generated reasoning ask the model to provide hints, questions, counterexamples, comparisons, or feedback on the learner's own reasoning. Examples include:

- “Do not give me the answer; give me one hint,”
- “Ask me one question at a time until I reach the solution,”
- “Find the weakest assumption in my argument,”
- “What evidence would contradict my conclusion?”,
- “Give me the strongest objection to this position,”
- “Compare my method with a different method,” and
- “Show where my reasoning first becomes invalid.”

These are proposed design practices rather than claim that every such prompt has independently demonstrated superior learning outcomes.

This dialogic use is consistent with the difference between the unrestricted and guarded AI conditions in Bastani et al. [8]. In that high-school mathematics experiment, the guarded tutor largely mitigated the later learning penalty observed with unrestricted GPT access. The study therefore provides direct evidence that interface constraints can matter in at least one instructional context. It does not establish that every hint-based or Socratic AI tutor will produce the same effect across subjects, ages, or durations of use.

AI can also serve as an intellectual adversary. After a learner develops a position, the model can be instructed to challenge it from a competing theoretical perspective, identify unexamined assumptions, propose counterexamples, or ask for stronger evidence. This role reverses the usual direction of delegation. Instead of asking AI to think in place of the learner, the learner asks AI to create resistance against which thinking must improve.

### 8.3. Think Again Without AI: Retrieve, Reconstruct, And Transfer

The final phase is crucial within the proposed framework because it tests whether assistance has become independently accessible knowledge. After interacting with AI, the learner closes the tool and reconstructs the key content independently. The learner can explain the concept aloud, write the main argument from memory, answer questions without assistance, solve a new problem, teach the idea to another person, or apply the principle in a novel setting. Retrieval research

supports the use of unaided recall as a learning activity [5], [6], while the learning-performance distinction cautions against inferring durable competence from assisted practice alone [1].

This phase also changes the role of AI in assessment. The system can help create practice questions, but at least some retrieval and performance checks should occur without the system when the purpose is to estimate independent competence. Bastani et al. [8] demonstrate why this distinction matters: performance while AI was available did not predict equally strong performance after it was removed. AI may therefore support rehearsal, but strategically placed unaided assessment can serve as a diagnostic of what the learner can retrieve or perform independently.

## 9. WHAT SHOULD BE OUTSOURCED - AND WHAT SHOULD NOT

The framework can be summarized by a task-specific principle: automate the peripheral, preserve the essential. The difficulty is that "peripheral" and "essential" are defined by the learning objective, not by the technology. Spell-checking may be peripheral in a history essay but central in a spelling lesson. Generating code documentation may be peripheral in a course on algorithms but central in a course on technical communication. A calculator may be appropriate in statistical inference but inappropriate when the objective is arithmetic fluency.

Several forms of AI use should therefore be treated as higher-risk when the goal is learning. In high-school mathematics, unrestricted answer-oriented GPT access harmed later independent performance relative to no AI, whereas a guarded tutor largely mitigated the effect [8]. More generally, requesting a complete answer before making an attempt removes an opportunity to generate a response, and lightly editing AI-generated prose offers less direct practice in constructing an argument than writing one. If summaries systematically replace assigned long-form reading, learners receive less practice in sustained reading and following extended arguments; this is a pedagogical inference rather than a demonstrated universal effect. Accepting citations or claims without verification is especially risky because fabricated and erroneous references have been empirically documented [15], and the over-reliance literature emphasizes the problem of uncritical acceptance of AI output [12].

Conversely, AI can be highly valuable when it reduces nonessential friction while leaving the target cognitive process intact. It can define an unfamiliar term, translate a sentence, reorganize notes, create additional practice items, diagnose an error after an attempt, propose counterarguments, simulate an oral examiner, generate questions for retrieval practice, or adapt the complexity of an explanation. The distinction is not between "AI use" and "no AI use," but between substitution and scaffolding.

## 10. FROM AI LITERACY TO COGNITIVE AI LITERACY

AI literacy frameworks have emphasized recognizing and understanding AI, using and applying it, evaluating outputs, creating with AI, and navigating ethical implications [18]-[20]. These dimensions remain essential. For generative AI specifically, learners should also understand that fluent output can be inaccurate or unsupported [12], [15], and that bias, privacy, authorship, and ethical use require active human judgment [18]-[20]. Yet the spread of highly capable generative systems adds another dimension: learners must understand how AI use can alter their own cognitive behavior.

This article therefore proposes cognitive AI literacy as a complementary component of AI literacy. Cognitive AI literacy is the ability to recognize which mental processes are being

delegated, predict when delegation may interfere with learning, and deliberately structure AI use so that target cognitive capabilities continue to be exercised. It includes knowing when not to ask for an answer, when to attempt retrieval without support, when to verify a claim independently, and when to choose a slower process because the effort itself is educationally productive.

A cognitively literate AI user is not simply the person who can obtain the best output from a model. It is the person who can decide which output should not be requested because receiving it would bypass the capability being developed. This is a more demanding form of digital competence because it requires metacognitive awareness of both the tool and the self.

## 11. IMPLICATIONS FOR TEACHING, ASSESSMENT, AND EDUCATIONAL POLICY

Educational institutions should avoid treating AI policy as a binary choice between prohibition and unrestricted permission. Both extremes ignore the interaction between tool design and learning design. The more useful question is whether a particular assignment requires students to expose and exercise the cognitive process being assessed. If a task can be completed entirely by an AI system without the student understanding the output, then the assessment may no longer measure the intended competence.

One practical design is a three-record assignment. Students submit an initial response created without AI, a record or summary of their interaction with AI, and a final response explaining what changed and why. The initial response reveals prior reasoning, the interaction shows how the tool was used, and the final response requires reflection and justification. Teachers can also ask students to identify errors in AI-generated answers, compare competing AI responses, locate primary sources that support or contradict a generated claim, or defend their written work orally. Such tasks shift evaluation from product alone to the evolution of reasoning.

Assessment should also include strategically placed AI-free components. Short oral defenses, handwritten or supervised problem solving, retrieval quizzes, live coding explanations, and in-class argument analysis can provide evidence of independent competence. The objective is not surveillance for its own sake but valid measurement. If every assessed performance is produced with continuous AI assistance, educators lose information about what students themselves know and can do.

Teachers should explicitly model cognitively productive prompts. Students need examples of how to ask for hints rather than answers, how to request criticism rather than completion, and how to verify model outputs. Families can reinforce similar norms by treating AI as a tutor that asks questions and provides explanations rather than as a homework-completion service. At the institutional level, AI literacy initiatives should include cognitive offloading, metacognitive calibration, verification habits, and the distinction between assisted performance and independent learning - not only technical prompting and academic-integrity rules.

Younger learners may benefit from greater external structure because at least some forms of metacognitive calibration continue to develop across adolescence. Weil et al. [21], using a perceptual decision task with participants aged 11 to 41 years, found that metacognitive ability improved during adolescence and plateaued toward adulthood. This finding concerns a specific form of metacognition and should not be generalized to every aspect of adolescent self-regulation. It nevertheless supports caution against assuming adult-level metacognitive monitoring in all younger learners. Learning environments can therefore build productive

structure into AI use through required first attempts, delayed answer reveal, hint sequences, source-check steps, and independent reconstruction after assistance.

## 12. PRODUCTIVE COGNITIVE FRICTION

The concept of productive cognitive friction captures a final principle. Efficiency is not always the correct metric for learning. Remembering without looking, attempting a problem before seeing the solution, discovering an error, revising an argument, and defending a conclusion can feel slower than accepting a fluent AI response. Yet research on the learning-performance distinction, retrieval practice, and generative learning shows that conditions requiring retrieval, generation, explanation, and transfer can support more durable learning than conditions that merely make practice feel fluent [1], [5]-[7]. Productive cognitive friction is therefore a design principle derived from established learning science, not a claim that difficulty is always beneficial.

The best educational AI may therefore be a system that sometimes refuses to be maximally convenient. It may ask what the learner has tried, provide a partial cue, request a justification, challenge an assumption, or withhold a complete solution until the learner has engaged the problem. Such friction is not a defect when the objective is learning. It is a design feature that keeps cognition active.

## 13. LIMITATIONS AND RESEARCH AGENDA

The framework proposed here is grounded in established learning science and emerging AI research, but the evidence base has important limitations. Longitudinal studies of generative AI use across childhood, adolescence, and higher education remain scarce. The strongest causal evidence cited here is task-specific: Bastani et al. [8] studied high-school mathematics. The two studies linking AI dependency and critical thinking are cross-sectional [10], [11], and the EEG essay-writing study remains a preprint [16] with published methodological criticism [17]. Effects observed in mathematics, survey measures, or essay writing may not generalize directly to other domains. Individual differences in prior knowledge, motivation, metacognitive skill, socioeconomic access, disability, and teacher support may also influence whether AI functions as scaffold or substitute.

Future research should therefore compare interaction patterns rather than merely AI access. Studies should distinguish unrestricted answer generation from hint-based tutoring, critique, retrieval support, and guided self-explanation. Outcomes should include delayed retention, transfer, calibration of confidence, independent writing and problem solving, source verification, and the ability to detect AI errors. Longitudinal research should examine whether habitual delegation changes the frequency with which learners initiate effortful reasoning outside AI-supported contexts. Developmental research is especially important because the optimal balance between external scaffolding and learner autonomy may vary substantially by age and expertise.

## 14. CONCLUSIONS

Generative artificial intelligence does not inevitably weaken or strengthen cognition. The evidence reviewed here supports a more conditional claim: educational outcomes depend on what the learner delegates, how the interaction is structured, the task being learned, and whether independent cognitive work is preserved before and after assistance. The clearest demonstrated risk in this literature is task-specific rather than universal-unrestricted answer-oriented assistance impaired later unaided performance in one large high-school mathematics experiment [8].

Broader concerns about dependency and critical thinking are presently supported mainly by cross-sectional associations [10], [11] and preliminary evidence [16], not by evidence of generalized cognitive decline.

Cognitive load theory provides a rationale for why AI may be beneficial when it removes demands that are extraneous to a learning objective, while also cautioning that working-memory relief is not equivalent to learning [3], [4]. Cognitive offloading explains why external assistance is a normal part of cognition [2]; the educational risk identified here arises when the offloaded operation is itself the capability that needs practice. Retrieval and generative learning research supports reconstructing, explaining, and applying knowledge rather than relying only on recognition of a fluent answer [5]-[7]. In generative-AI research, Bastani et al. [8] provide the strongest causal evidence cited here, while dependence studies [10], [11] and the preliminary EEG study [16], [17] should be interpreted according to their more limited designs.

The resulting principle is concise: think before AI, think with AI, and think again without AI. Used in this sequence, generative AI can function as scaffold, tutor, critic, and intellectual adversary. Used as a permanent substitute for unfinished cognitive capabilities, it can make the product appear stronger than the learner has become. Education in the age of generative AI should therefore aim not simply to teach students how to obtain answers from intelligent systems, but how to use those systems while preserving the capacity to reason when the systems are absent, wrong, or unable to decide what matters.

## 15. STATEMENTS AND DECLARATIONS

### 15.1. Conflict of Interest

The author declares no known financial or non-financial competing interests related to this manuscript.

### 15.2. Funding

No external funding was received for the preparation of this conceptual article.

### 15.3. Data Availability

No original datasets were generated or analyzed because this article is a conceptual review of published literature.

### 15.4. Ethical Statement

This manuscript does not report research involving human participants, animals, or identifiable personal data; therefore, institutional ethics approval and informed consent were not applicable.

### 15.5. Declaration of Generative AI Use

During preparation of this manuscript, the author used OpenAI ChatGPT (GPT-5.6 Sol; accessed August 28, 2026) to translate and edit an author-originated Portuguese manuscript and to improve academic English, readability, clarity, and organization. The substantive ideas, central framework, interpretations, and conclusions originated with and were approved by the author. Bibliographic metadata and claim-to-source alignment were checked against publisher records, DOI-linked academic databases, and academic indexing services. The author reviewed the AI-

assisted text and accepts full responsibility for the final content. Generative AI was not used to collect or generate research data, perform statistical analyses, determine study results, or create scientific figures or images.

## REFERENCES


[1] Soderstrom, N. C. & Bjork, R. A. (2015) "Learning versus performance: An integrative review," Perspectives on Psychological Science, Vol. 10, No. 2, pp. 176-199. https://doi.org/10.1177/1745691615569000

[2] Risko, E. F. & Gilbert, S. J. (2016) "Cognitive offloading," Trends in Cognitive Sciences, Vol. 20, No. 9, pp. 676-688. https://doi.org/10.1016/j.tics.2016.07.002

[3] Sweller, J. (1988) "Cognitive load during problem solving: Effects on learning," Cognitive Science, Vol. 12, No. 2, pp. 257-285. https://doi.org/10.1207/s15516709cog1202_4

[4] Sweller, J., van Merriënboer, J. J. G. & Paas, F. (2019) "Cognitive architecture and instructional design: 20 years later," Educational Psychology Review, Vol. 31, No. 2, pp. 261-292. https://doi.org/10.1007/s10648-019-09465-5

[5] Karpicke, J. D. & Blunt, J. R. (2011) "Retrieval practice produces more learning than elaborative studying with concept mapping," Science, Vol. 331, No. 6018, pp. 772-775. https://doi.org/10.1126/science.1199327

[6] Roediger, H. L., III & Butler, A. C. (2011) "The critical role of retrieval practice in long-term retention," Trends in Cognitive Sciences, Vol. 15, No. 1, pp. 20-27. https://doi.org/10.1016/j.tics.2010.09.003

[7] Fiorella, L. & Mayer, R. E. (2016) "Eight ways to promote generative learning," Educational Psychology Review, Vol. 28, No. 4, pp. 717-741. https://doi.org/10.1007/s10648-015-9348-9

[8] 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 of the United States of America, Vol. 122, No. 26, e2422633122. https://doi.org/10.1073/pnas.2422633122

[9] Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S, M. & Joseph, S. (2025) "The cognitive paradox of AI in education: Between enhancement and erosion," Frontiers in Psychology, Vol. 16, 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

[10] Tian, J. & Zhang, R. (2025) "Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy," Acta Psychologica, Vol. 260, 105725. https://doi.org/10.1016/j.actpsy.2025.105725 Corrigendum: Tian, J. & Zhang, R. (2026) "Corrigendum to 'Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy'," Acta Psychologica, Vol. 267, 107179. https://doi.org/10.1016/j.actpsy.2026.107179

[11] Tian, J. & Zhang, R. (2026) "Outsourcing thinking to AI? Focused immersion, AI dependency, and the double-edged impact on critical thinking," Humanities and Social Sciences Communications, Vol. 13, 1015. https://doi.org/10.1057/s41599-026-07153-8

[12] Zhai, C., Wibowo, S. & Li, L. D. (2024) "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review," Smart Learning Environments, Vol. 11, 28. https://doi.org/10.1186/s40561-024-00316-7

[13] Sparrow, B., Liu, J. & Wegner, D. M. (2011) "Google effects on memory: Cognitive consequences of having information at our fingertips," Science, Vol. 333, No. 6043, pp. 776-778. https://doi.org/10.1126/science.1207745

[14] 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, Vol. 14, No. 1, pp. 4-58. https://doi.org/10.1177/1529100612453266

[15] Walters, W. H. & Wilder, E. I. (2023) "Fabrication and errors in the bibliographic citations generated by ChatGPT," Scientific Reports, Vol. 13, 14045. https://doi.org/10.1038/s41598-023-41032-5

[16] Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I. & Maes, P. (2025) "Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task," arXiv preprint, arXiv:2506.08872, revised December 31, 2025. https://doi.org/10.48550/arXiv.2506.08872

[17] Stanković, M., Hirche, E., Kollatzsch, S. & Doetsch, J. N. (2025) "Comment on: Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Tasks," arXiv preprint, arXiv:2601.00856. https://doi.org/10.48550/arXiv.2601.00856

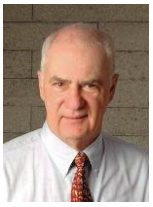
[18] Almatrafi, O., Johri, A. & Lee, H. (2024) "A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019-2023)," Computers and Education Open, Vol. 6, 100173. https://doi.org/10.1016/j.caeo.2024.100173

[19] Long, D. & Magerko, B. (2020) "What is AI literacy? Competencies and design considerations," 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

[20] Ng, D. T. K., Leung, J. K. L., Chu, K. W. S. & Qiao, M. S. (2021) "AI literacy: Definition, teaching, evaluation and ethical issues," Proceedings of the Association for Information Science and Technology, Vol. 58, No. 1, pp. 504-509. https://doi.org/10.1002/pra2.487

[21] Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J. & Blakemore, S.-J. (2013) "The development of metacognitive ability in adolescence," Consciousness and Cognition, Vol. 22, No. 1, pp. 264-271. https://doi.org/10.1016/j.concog.2013.01.004

## AUTHOR



**Dr. Ronald Jean Degen** is a Swiss academic living in Brazil, author, and business executive widely recognized as the pioneer of entrepreneurship education in Brazil. With a career spanning over four decades, he has successfully bridged the gap between academic theory and corporate practice.

- Credentials: He earned a Ph.D. in International Business from the International School of Management (ISM) in Paris, MBA from the University of Michigan in Ann Arbor, Pos Graduate Degree in Automation from the ETH in Zurich, an Electronic Engineer Degree from the Istituto Mauá in São Paulo
- Research Contributions: His extensive academic research focuses heavily on strategic decision-making, executive leadership, business ethics, innovation capabilities, and shifting frameworks from shareholder to stakeholder.
  https://ronaldjeandegen.academia.edu/research
- **Executive Experience:** He has spent over 30 years as a CEO, board member, and turn-around specialist across diverse international markets.
- **Major Roles:** He previously served as the Vice Chairman of Masisa in Chile and has successfully led corporate growth, mergers and acquisitions, and strategic repositioning for multiple enterprises.
- **Credentials:** He earned a Ph.D. in International Business from the International School of Management (ISM) in Paris, MBA from the University of Michigan in Ann Arbor, Pos Graduate Degree in Automation from the ETH in Zurich, an Electronic Engineer Degree from the Istituto Mauá in São Paulo
- **Research Contributions:** His extensive academic research focuses heavily on strategic decision-making, executive leadership, business ethics, innovation capabilities, and shifting frameworks from shareholder to stakeholder.
  https://ronaldjeandegen.academia.edu/research