# GLOBAL JOURNAL
OF COMPUTER SCIENCE AND TECHNOLOGY - G: INTERDISCIPLINARY

Volume 26 | Issue 1 | 2026 | Article ID: GJCST-257948 | Language: English

CATALOGING

ACM • K.3.1, H.1.2 | ARXIV • cs.HC, cs.CY | ANZSRC FOR • 3901


# Does Artificial Intelligence Make You Stupid? Cognitive Debt and a Three-Tier Framework for Sustainable Human–AI Collaboration

Article Record


**Khaled EL Tannir**[§‡ϑ*]

Founder & CEO

*Corresponding Author

§ dataXper, Montréal, Canada
‡ McGill University, Montreal, Canada (**OA**)



| RECEIVED | ACCEPTED | ONLINE PUBLISHED | PUBLISHED | PEER REVIEW |
|---|---|---|---|---|
| 2026-06-03 | 2026-06-13 | 2026-06-26 | 2026-07-10 | Double Blind |



## Abstract

Generative artificial intelligence now drafts our correspondence, condenses our reading, and solves our problems faster than we can articulate them. A growing body of evidence suggests, however, that the very tools that lift short-term output can quietly erode the cognitive capacities they appear to augment. This paper synthesizes four recent empirical studies—a randomized controlled trial of AI in high-school mathematics, a field experiment with elite management consultants, a classroom trial of an AI physics tutor, and a neurophysiological study of AIassisted writing—and argues that the headline question is badly posed. The decisive variable is not whether people use AI but how they use it. We adopt cognitive debt, a construct describing the trade of present convenience for future capability, as an organizing lens, and develop a threetier framework for cognitively sustainable use. Tier 1 delegates non-developmental tasks without guilt; Tier 2 protects domain expertise by deploying AI adversarially rather than substitutively; Tier 3 preserves the desirable difficulties on which durable learning depends. Each tier is grounded in established findings from cognitive psychology. We close with implications for individuals and organizations, and with the limitations of an evidence base that remains early and uneven.



generative AI | cognitive offloading | cognitive debt | learning | expertise | human–AI collaboration | desirable difficulties

**AI USE STATEMENT**
No generative AI was used for analysis or results.

**FUNDING**
No external funding was declared for this work.

**CONFLICT OF INTEREST**
The authors declare no conflict of interest.

**DATA AVAILABILITY**
Not applicable for this article.

**ETHICS**
No ethics committee approval was required for this article type.

**CONSENT**
Not applicable for this article.

**TRIAL REG.**
Not applicable.

**Crossref DOI:** 10.34257/GJCSTG257948

**How to Cite:** EL Tannir (2026). Does Artificial Intelligence Make You Stupid? Cognitive Debt and a Three-Tier Framework for Sustainable Human–AI Collaboration. Global Journal of Computer Science and Technology, 26(1), 1-4. DOI: 10.34257/GJCSTG257948





AR Experience Web Page DOI

Print ISSN 0975-4350
9 770975 435008

Online ISSN 0975-4172
9 770975 417011

*Under the strict compliance and defined process of*

## METADATA CONTINUATION

### AUTHOR CONTACT QR LEDGER

Khaled EL Tannir§‡ϑ*


### ARCHIVAL RECORD

GJCST · Vol 26 · Issue 1 · 2026

Article ID GJCST-257948 · DOI 10.34257/GJCSTG257948

Print ISSN 0975-4350 · Online ISSN 0975-4172

Global Journal of Computer Science and Technology - Section G
GJCST Volume 26 Issue 1

# Does Artificial Intelligence Make You Stupid? Cognitive Debt and a Three-Tier Framework for Sustainable Human–AI Collaboration

Khaled EL Tannir[§‡ϑ*]

**Affiliations**
§ dataXper, Montréal, Canada
‡ McGill University, Montreal, Canada (OA)

**Qualifications / Designations**
ϑ Founder & CEO

**Abstract**

Generative artificial intelligence now drafts our correspondence, condenses our reading, and solves our problems faster than we can articulate them. A growing body of evidence suggests, however, that the very tools that lift short-term output can quietly erode the cognitive capacities they appear to augment. This paper synthesizes four recent empirical studies—a randomized controlled trial of AI in high-school mathematics, a field experiment with elite management consultants, a classroom trial of an AI physics tutor, and a neurophysiological study of AIassisted writing—and argues that the headline question is badly posed. The decisive variable is not whether people use AI but how they use it. We adopt cognitive debt, a construct describing the trade of present convenience for future capability, as an organizing lens, and develop a threetier framework for cognitively sustainable use. Tier 1 delegates non-developmental tasks without guilt; Tier 2 protects domain expertise by deploying AI adversarially rather than substitutively; Tier 3 preserves the desirable difficulties on which durable learning depends. Each tier is grounded in established findings from cognitive psychology. We close with implications for individuals and organizations, and with the limitations of an evidence base that remains early and uneven.

**Keywords:** *generative AI, cognitive offloading, cognitive debt, learning, expertise, human–AI collaboration, desirable difficulties*

*** Corresponding Author**
Khaled EL Tannir

**DOI**
10.34257/GJCSTG257948

## 1. Introduction

Consider a result that ought to unsettle anyone who reaches for a chatbot before they reach for their own judgment. In a large randomized controlled trial conducted in a Turkish high school, researchers affiliated with the University of Pennsylvania and the Wharton School gave nearly a thousand mathematics students access to a GPT-4-based assistant during practice sessions (Bastani et al., 2025). While the assistant was available, performance soared: students working with a standard chatbot interface solved practice problems roughly forty-eight percent better than peers with no such help. The intervention looked like an unambiguous success—until the assistant was taken away for the final, unaided examination. Stripped of the tool, the students who had leaned on it scored measurably below classmates who had never used it at all. The crutch, it turned out, had been doing the walking.

This pattern is not confined to classrooms, and it is the central concern of this paper. The same logic——impressive output while the tool is present, diminished capability once it is withdrawn—appears wherever cognitive work is silently transferred from a person to a machine. The popular question, posed bluntly, is whether AI makes us stupid. We will argue that this framing is unhelpful, because it treats a tool as if it had a single fixed effect on its user. It does not. A hammer neither builds nor destroys a house on its own; what matters is the hand that holds it and the purpose to which it is put. The relevant question is therefore not whether one uses AI, but how.

The remainder of the paper proceeds in four movements. Section 2 assembles the empirical case for concern, drawing on four recent studies that span education, professional knowledge work, and neurophysiology. Section 3 reframes the problem around the construct of cognitive debt and explains why a how-question is more tractable and more honest than a whether-question. Section 4 presents the paper's main contribution: a three-tier framework—delegation, augmentation, and cognitively active learning—each tier anchored in well-established results from cognitive psychology. Section 5 discusses implications and limitations, and Section 6 concludes.

## 2. Productivity without Proficiency: the Empirical case

Three findings recur across otherwise dissimilar studies. AI reliably raises immediate output. That output tends to converge toward sameness across users. And the gains often fail to survive the tool's removal, leaving the user worse off than if they had struggled alone. Taken together, these findings describe a dissociation between productivity and proficiency that the headline figures of the AI era tend to obscure.

### 2.1. Learning: the crutch effect

The Bastani et al. (2025) trial, published in the Proceedings of the National Academy of Sciences, is among the most carefully designed studies of the question to date. Students were randomly assigned to one of three conditions during their practice sessions:

an unconstrained chatbot built on GPT-4 ("GPT Base"), a pedagogically constrained version designed with teacher input to offer hints rather than answers ("GPT Tutor"), or no AI access at all. During assisted practice, both AI groups outperformed the control—the constrained tutor dramatically so. But on a subsequent unaided examination, the students who had used the unconstrained chatbot performed worse than those who had never touched it. The authors interpret this as evidence that an answerdispensing assistant functions as a crutch: it allows students to produce correct work without performing the cognitive operations that produce learning.

The design also contains the seed of an antidote, a point we return to in Section 4.3. The constrained GPT Tutor group, whose assistant withheld direct answers and instead guided students toward solutions, suffered no comparable penalty on the unaided exam; their scores were statistically indistinguishable from the control group. The same underlying model, configured to question rather than to answer, neutralized the harm. The lesson is not that AI tutoring is dangerous in itself, but that an interface optimized to be maximally helpful in the moment can be precisely the wrong thing for learning over time.

### 2.2. Expertise: better work, fewer ideas

If the classroom shows what AI does to the acquisition of skill, a landmark field experiment shows what it does to the exercise of expertise. Dell'Acqua and colleagues (2023), working with the Boston Consulting Group and a team drawn from Harvard, Wharton, and MIT, randomly assigned 758 management consultants—roughly seven percent of the firm's individual contributors—to complete realistic consulting tasks with or without GPT-4. On tasks lying within the model's competence, the effects were large: consultants using AI completed more tasks, finished them faster, and produced work judged about forty percent higher in quality. This is the result that made headlines.

The result that did not make headlines is, for our purposes, more important. The firm's own analysis of the experiment found that the relatively uniform output of the model reduced the group's diversity of thought by roughly forty-one percent. Individually, almost everyone improved; collectively, they converged. Consultants drew on similar framings, advanced similar arguments, and arrived at similar conclusions. The study also surfaced a "jagged technological frontier": on tasks lying just outside the model's competence, AI-assisted consultants were markedly more likely to be wrong, in part because they extended to the machine a trust it had not earned. Expertise has always consisted in part of idiosyncrasy—the particular way a seasoned practitioner sees a problem that others miss. A homogenizing tool puts exactly that idiosyncrasy at risk, and with it the differentiated judgment that gives an expert market value. The same compression of novel output has since been observed in creative writing (Doshi & Hauser, 2024).

### 2.3. Cognition: the neural signature of offloading

A third study moves the question from behavior to the brain. In a preprint from the MIT Media Lab, Kosmyna and colleagues (2025) asked fifty-four participants to write essays under one of three conditions—using a large language model, using a search engine, or using no external tool at all—while their brain activity was recorded with electroencephalography. The participants who wrote with the language model showed the weakest and least distributed neural connectivity of the three groups; the unaided writers showed the strongest. The model-assisted writers also reported a diminished sense of ownership over what they had produced and were less able to recall or quote their own essays moments after finishing them. Notably, when participants who had relied on the model were later asked to write without it, the under-engagement persisted. The authors named this accumulated cost cognitive debt: a deferral of mental effort that, like a financial debt, is convenient now and expensive later.

Two caveats are essential and we state them plainly, because an academic reader will check. The study is a preprint and has not, at the time of writing, completed peer review; its sample is small and its task is narrow. The authors themselves caution against the sensational reading—they did not measure intelligence and explicitly resist the language of "brain rot." The construct of cognitive debt is valuable less as a proven neurological fact than as a precise name for a mechanism that the behavioral studies independently imply

## 3. From "Whether" to "How" ": Reframing the Question

The mechanism underlying all three findings is well understood and considerably older than generative AI. Psychologists call it cognitive offloading: the use of external aids to reduce the mental demands of a task (Risko & Gilbert, 2016). Offloading is not inherently harmful—writing offloads memory, and few would surrender it—but it carries a consistent cost. When we delegate the act of retrieval, we weaken the memory that retrieval would have built. The point was demonstrated for internet search well before chatbots existed: people who expect to have continued access to information online remember the information itself less well, while remembering where to find it (Sparrow et al., 2011). A language model is a cognitive offloading device of unprecedented reach, capable of absorbing not merely storage but reasoning, composition, and judgment. The more it absorbs, the more of those faculties go unexercised.

This reframing matters because it changes what we are permitted to conclude. The studies in Section 2 do not show that AI makes its users less capable. They show that a particular mode of use—substituting the tool for the cognitive work rather than supporting it—produces that result and that a different mode of use does not. The unconstrained chatbot harmed learning; the Socratic tutor did not. The consultants who let the model think for them converged and erred at the frontier; the construct that predicts who suffers cognitive debt is the depth of substitution, not the mere presence of the tool. The honest question, then, is a design question, addressed to the user as much as to the engineer: which cognitive operations should be handed over, and which must remain our own? The next section offers a structured answer.

## 4. A Three-Tier framework for Cognitively Sustainable use

We propose that AI use be sorted into three tiers according to a single criterion: the relationship between the task and the user's own development. Where a task demands no skill the user needs to build, full delegation is appropriate and even obligatory. Where a task lies within the user's domain of expertise, the tool should augment judgment without replacing it. Where the explicit goal is to learn, the tool must be configured to preserve effort rather than remove it. The tiers are not a hierarchy of value but a partition of situations; a single person moves among all three within a working day.

### 4.1. Tier 1 — Delegation: handing over what does not develop you

A great deal of knowledge work consists of tasks for which the user has neither expertise nor any reason to acquire it. Reformatting a dataset, transcribing a recording, translating boilerplate, producing a serviceable first draft, or synthesizing fifteen sources into a brief—these are the cognitive equivalent of errands. Delegating them to AI is not a moral failing but a reallocation of scarce attention toward the work that genuinely requires it. The analogy is to any competent professional one engages: one does not re-audit every line an accountant prepares, because the point of the arrangement is to be relieved of that labor.

Four categories delegate cleanly. First, the blank-page draft: the value of a generated first draft lies less in its words than in the momentum it supplies, after which the user's own judgment takes over to revise. Second, broad research synthesis, where the tool compresses an afternoon of reading into minutes. Third, pattern detection across large bodies of data—sales figures, customer feedback—where a model surfaces regularities a person could not hold in mind at once. Fourth, mechanical transformation: cleaning, reformatting, transcription, translation. The single discipline that Tier 1 requires is verification. Because models are confidently wrong on a non-trivial fraction of outputs—the jagged frontier of Section 2.2—delegation is not abdication. A human eye on the result before it leaves your hands is the price of admission, and it is cheap.

### 4.2. Tier 2 — Augmentation: protecting expertise

The second tier is where the most consequential and least visible damage occurs. Within one's own field, the temptation is to use AI exactly as in Tier 1—to let it produce the analysis, the argument, the recommendation—and the immediate output is often excellent. But this is precisely the substitution that the consulting experiment showed to homogenize thought and the writing study showed to incur cognitive debt. The faculty being offloaded here is not clerical; it is the differentiated judgment that constitutes expertise itself.

The governing image for Tier 2 is the athletic coach. A coach never runs the race in the athlete's place. The coach observes, challenges, asks the uncomfortable question, and pushes the athlete past the point they would have stopped at alone—but the exertion, and the resulting strength, belong to the athlete. Used this way, AI is asked not to produce the work but to interrogate it: to find the flaw in an argument, to supply the strongest objection, to propose an angle the expert has not considered. The analysis, the weighing, and the final decision remain the human's. The distinction is subtle in practice and decisive in consequence. Asking a model "write my recommendation" offloads the judgment; asking it "here is my recommendation—where is it weakest?" exercises the judgment against resistance. The first accrues cognitive debt; the second pays it down.

### 4.3. Tier 3 — Cognitively Active learning: keeping the difficulty

The third tier addresses the case the popular discourse most often gets wrong: using AI to learn. The intuitive approach—"summarize this chapter," "make me a study sheet," "explain this so I don't have to read it"—delegates the cognitive labor of learning to the machine, which is to say it delegates the learning. This is the mechanism behind the examination collapse in Bastani et al. (2025): the chatbot did the work, and the work was the point.

Cognitive psychology has spent decades establishing why this fails, and the findings cohere into a single principle: effortful processing is what makes learning durable. Bjork and Bjork (2011) describe desirable difficulties—conditions that slow acquisition and feel harder in the moment but markedly improve long-term retention and transfer. The generation effect shows that information a learner produces is remembered better than information merely read (Slamecka & Graf, 1978).

The testing efect shows that retrieving knowledge through self-testing strengthens memory more than re-studying does (Roediger & Karpicke, 2006). And cognitive load theory explains why a tutor that hands over answers can hurt: it removes the germane load—the productive effort of building a mental schema—that learning requires (Sweller, 1988). Each of these effects is abolished when an AI removes the effort.

The corrective is to configure the tool to impose desirable difficulty rather than relieve it—to make it a Socratic tutor rather than an oracle. In practice this means inverting the usual prompts. Instead of "summarize this concept," ask the model to test you: "ask me questions to check whether I have understood this," "build me a quiz of increasing difficulty," "challenge my answers and tell me where my reasoning breaks." The constrained tutor in Bastani et al. (2025), which guided with hints rather than answers, did no harm to unaided performance; and a Harvard trial of a purposebuilt physics tutor designed along these lines found that students learned more than twice as much, in less time, as peers in a well-run active-learning classroom (Kestin et al., 2025). The published evidence appeared in Scientific Reports, a Nature Portfolio journal—not, as it is sometimes reported, in Nature itself—and rests on fewer than two hundred students, so it should be read as promising rather than definitive. Its mechanism, however, is exactly the one the broader literature predicts: a tutor that preserves effort preserves learning. A useful side benefit, often noted by users, is that one can ask a machine the most elementary question without embarrassment—lowering the social cost of difficulty while keeping the cognitive cost intact.

## 5. Discussion

The three tiers share one accounting principle. Cognitive debt is incurred whenever a person offloads a cognitive operation they would otherwise have needed to perform themselves for their own development, and it is avoided—or repaid—whenever the tool is arranged to demand effort rather than absorb it. Tier 1 incurs no debt because the offloaded operations were never developmental. Tier 2 risks the most insidious debt, because the offloading feels productive and the loss—of original judgment—is invisible until the tool is unavailable or wrong. Tier 3 converts the tool from a creditor into a trainer.

At the level of the organization, the homogenization finding deserves particular attention. If every analyst, every firm, and every competitor consults the same handful of models, the convergence documented among the consultants scales into a market-wide flattening of ideas.

Differentiation—the source of competitive advantage and of intellectual progress alike—depends on variance, and variance is what a shared statistical model erodes. The individual incentive to use AI for higher immediate quality and the collective interest in a diversity of thought can therefore point in opposite directions, a tension organizations would do well to manage explicitly rather than discover after the fact.

Several limitations bound these conclusions. The evidence base is young and uneven: the strongest learning study is a single-country school trial, the consulting experiment is confined to one elite firm and one model generation, the physics trial is small, and the neurophysiological study is an unreviewed preprint with a narrow task. The capabilities of the models themselves are a moving target; the jagged frontier of 2023 is not the frontier of today, and findings tied to GPT-4 may not transfer cleanly to later systems. Cognitive debt, finally, remains a construct rather than a settled quantity—useful for organizing the evidence and guiding practice, but not yet a measured variable with an agreed unit. None of these caveats overturns the central, convergent observation: across education, professional work, and the laboratory, the same tool produces opposite effects depending on whether it is allowed to replace cognition or required to provoke it.

### 5.1. Implications for practice

For the individual—student, researcher, or professional—the operative discipline is triage before delegation. Before opening a chatbot, ask which tier the task occupies: whether the work is something you need to be able to do, or merely something that needs to be done. A literature search synthesized by a model is a Tier 1 errand; the argument of your own paper is not, and outsourcing it forfeits the very judgment a research career is built upon. Learners should treat the "summarize this for me" reflex as a warning sign and invert it, asking the tool to test them rather than to tell them, since the effort it removes is precisely the effort that consolidates knowledge (Bastani et al., 2025; Kestin et al., 2025).

For practitioners and managers, the homogenization result carries the sharper warning. If a team's analysts all consult the same model, the convergence Dell'Acqua and colleagues (2023) observed will quietly erase the differentiation on which competitive advantage depends. The remedy is procedural: have people commit their own framing to paper before they consult AI, use the tool to stress-test conclusions rather than to generate them, and treat any task near the jagged frontier as one requiring human verification rather than trust. Organizations should also resist measuring AI's value by immediate output alone. A metric that rewards faster, higher-quality deliverables while ignoring the slow erosion of in-house expertise will optimize for exactly the cognitive debt this paper warns against. In both cases the aim is not less AI but better-placed AI.

## 6. Conclusion

We live through a genuinely unusual moment, holding an instrument that could make us the most capable thinkers any generation has been—or could leave us unable to think without it. The evidence reviewed here suggests that both outcomes are available from the same device, and that the difference between them lies almost entirely in the manner of use. Delegate what does not develop you; guard the judgment that makes you an expert by using AI to challenge it rather than to supply it; and when the goal is to learn, keep the difficulty that learning requires. The question worth asking is not whether you use artificial intelligence. It is whether you use it to think better, or to avoid thinking at all. Adaptation, in the end, is not the act of using AI. It is the discipline of knowing how.

## REFERENCES


[1] Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakç1, 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).

[2] Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher et al. (Eds.), Psychology and the real world (pp. 5664). Worth Publishers.

[3] Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality (Harvard Business School Working Paper No. 24-013).

[4] Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28).

[5] Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms active learning. Scientific Reports, 15. [Nature Portfolio]

[6] 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:2506.08872) [Preprint].

[7] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676688.

[8] Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves longterm retention. Psychological Science, 17(3), 249255.

[9] Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, 4(6), 592604.

[10] Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science, 333(6043), 776778.

[11] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257285.