Overview of the MIT Study

In their study "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task", researchers at MIT Media Labs set out to examine how the use of a large language model (LLM), such as ChatGPT, might influence cognitive effort, neural engagement, learning, and a writer's sense of ownership over their work. The study asked whether AI-assisted writing changes how people think during writing tasks and whether these differences suggest reduced cognitive engagement over time.

Based on their findings, the authors argued that heavy reliance on AI tools may reduce critical thinking and memory formation by lowering cognitive effort when completing intellectual tasks. They concluded that AI support should be used cautiously in learning environments to avoid possible long-term erosion of deep cognitive habits.

Below is a critique of the study design and conclusions that may serve as a caution against assuming conclusions from this study regarding the benefits or consequences of AI use.

A Critique of "Your Brain on ChatGPT": Why the Study Falls Short

The MIT paper titled "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task" (https://arxiv.org/abs/2506.08872) presents itself as a sophisticated and credible exploration of how AI might alter cognitive functioning. At first glance, the work appears methodologically sound and technically impressive. It features EEG scans, discussions of connectivity in alpha and beta networks, careful p-value reporting, polished visualizations, and other impressive-sounding elements.

However, once you move past the visual and technical polish, the study suffers from foundational design flaws, interpretive overreach, and a mismatch between what was measured and what is claimed.

Descriptive, Not Causal: Limits of the Study's Claims

The work is best understood as descriptive and preliminary rather than definitive or causal. Descriptive research in emerging fields is valuable, but findings from such research cannot, and should not, be misconstrued as demonstrating causality. In this case, the study cannot claim that AI use *caused* deterioration of mental functioning.

This study describes what occurred during a set of controlled writing tasks. It does not establish causality, generalizability, cognitive change, or skill loss or brain degradation. Yet the authors write as if these conclusions naturally follow from the data. They do not.

The findings from this study are narrow and context-bound, and the paper draws causal implications not supported by its design. The authors propose a causal relationship between

cognitive ability and AI use (and certainly much of the attention this study has received is based on this misunderstanding), but that is simply not possible with this study.

What the Study Actually Measures Versus What It Claims

At most, the study shows that AI-assisted writing produces different observable effects than unaided writing—a far cry from proving that AI harms learning or cognition.

A central problem lies in the disconnect between performance on a task and cognitive capacity. The study analyzes essays, brain activity, and recall performance during a constrained writing exercise. This tells us only what participants did *under those specific task conditions*, not what they are cognitively capable of doing more broadly.

Reduced neural activation during AI-assisted writing neither demonstrates nor proves diminished cognitive ability. It simply indicates that the task required less effort when AI assistance was available. Importantly, students who used AI were able to complete the required tasks, regardless of any differences in cognitive effort.

In other words, this study observes behavior and activation, but it cannot infer mental capacity, degradation, or long-term change. The authors repeatedly imply or suggest otherwise, in spite of multiple alternative potential reasons for differences in observed effects.

Multiple plausible explanations exist for decreased neural activity, recall, and paper quality. In spite of the authors' argument that the study revealed harmful effects of AI use, the study, itself, does not isolate which explanation for any observed effect, if any, is correct.

Effort, Incentives, and Motivation: The Missing Variables

The study conditions created no reason for students to exert high cognitive effort. Participants received the same reward regardless of output quality or learning. The reward structure was flat-rate compensation for attendance, not for outcomes, engagement level, or quality of work. Participants were, therefore, incentivized only to complete the sessions, not to exert greater effort or retain information. That context is important when interpreting cognitive-effort and recall findings.

Under those conditions, it is rational to conserve effort, particularly when using AI tools that can automate portions of the task.

AI-using participants may therefore have been completing the task more efficiently or conserving cognitive resources. They may have prioritized minimal effort rather than deep learning. This is not evidence of impairment. It is evidence of strategy.

Motivation declines further across repeated sessions with no meaningful stakes. Participants may have recognized that "doing the minimum" was sufficient. This could reasonably explain why cognitive activation dropped over time: not degradation, but disengagement.

Memory and Recall: A Problem of Task Design, Not AI

One of the most publicized claims is that AI users could not recall their own essay content. However, the assigned task did not require retention, and the point of the task was not learning.

Participants were never told that recall would be assessed or rewarded. Their job was simply to produce text, which they accomplished as directed.

This means they had no purpose to encode the material into memory, and any recall failure may have reflected the task structure and assessment process rather than cognitive decline.

The study does highlight one legitimate instructional risk: students may not know AI-generated content unless the learning design requires them to do so. But that is not a flaw in AI itself; it is a flaw in assessment design. When learning tasks reward output rather than understanding, students logically optimize for output.

As I have described more fully in the EdAINow "Systems-Level Guide to Preventing Cheating with AI" (available at https://edainow.com/resources), well-designed AI-integrated learning environments address this by making students responsible for defending, applying, and reflecting on AI-assisted work. These factors were not present in this study design.

Contamination and Group Validity Concerns

In scientific experimental design, outcomes for a control group are compared to outcomes for a population that receives one or more "treatments". Members of the control group receive no treatment, which allows researchers to attribute differences in the outcomes to the treatment, or intervention.

This study uses the brain-only writing group as the control population. However, there is no evidence that these participants do not use AI in their regular academic or personal lives. Given the high usage rates self-reported by students, it is likely that many do. What this means is that the control group is probably not a separate population from the intervention population, and cannot actually serve as the control.

If the study's claims were true, then both groups' brains would already be "AI-affected." This invalidates the authors' premise. Any observed differences must, therefore, result from immediate task conditions, and not from long-term neurological impact.

Differences in participants' motivation, familiarity, attention, and engagement are much more plausible factors in the outcomes.

Interpreting Reduced Neural Activation: Efficiency and Motivation, Not Decline

The authors interpret declining neural connectivity in the AI-writing group as evidence of cognitive deterioration. But reduced mental effort could result from other factors. Students may have become more familiar with the task or more efficient with using AI to complete it. It would be reasonable to suggest that some participants may have simply disengaged from effort because the results did not have any personal impact, or because they became disinterested in the study.

Further, when brain-only writers switched to AI use in Session 4, their neural engagement increased. This alone challenges the authors' narrative that AI use suppresses neural involvement. If their interpretation were sound, the opposite should have occurred.

(On a side note: This would be an interesting phenomenon to study, as it relates to experiences of educators and others who are novices at using AI.)

Missing the Complexity of AI Use

The study treats all AI usage as equal, yet AI can be used at multiple levels of complexity:

- fully delegating a task,
- co-constructing ideas,
- seeking clarifications,
- iterating collaboratively, and
- synthesizing independently.

The task design and instructions appear to encourage simple delegation rather than meaningful synthesis. This artificially limits the cognitive engagement AI could potentially support. (For a more comprehensive discussion on levels of AI use, see our "AI Usage Guide for Instructional Design and Academic Integrity: The 5 Levels of AI Use for Students", available at https://edainow.com/resources.)

In real-world learning, the focus is not to replace thinking with AI-generated content, but to enhance thinking through guided, structured interaction. This study does not simulate that environment, and any results or observed effects cannot be generalized to broader contexts outside the study context.

Satisfaction, Ownership, and Human Psychology

The study finds that brain-only writers reported higher ownership and satisfaction, through the relationship to cognitive functioning is unclear. This is unsurprising and predictable. Effort justification theory, and simple lived experiences, indicate that people generally value things more when they invest significant effort. Lower satisfaction does not imply harm; it reflects natural and reasonable psychological valuation.

A reasonable explanation is that if the participants cannot answer "Why should I invest more than the minimum effort?", their level of effort will be lower, and, as a result, their satisfaction and ownership would logically be lower than participants who were required to work harder and invest more effort by the nature of the task.

Weaknesses in Research Design and Rigor

On top of the design flaws and questionable interpretation of the results, this study suffers from a number of credibility concerns.

First, the paper is not peer-reviewed, which is standard with scientific research publications. It is that process that lends credibility to results and identifies ideas and conclusions worthy of attention. Without peer review, there is no external validation of the results.

Second, fewer than 20 participants completed the final session, making longitudinal claims unreliable, to say the least. Simply put, there was not sufficient evidence to support conclusions or to generalize to a broader population.

Third, the researchers used AI in analysis while simultaneously suggesting AI usage impairs cognition. This is contradictory. If the researchers' claims are, in fact, true, then conclusions based on AI analysis would be suspect.

Fourth, the study uses research questions rather than null hypotheses, yet frames the results as confirmatory. This may be the major deterrent to any conclusions regarding causality. A null hypothesis is essential to experimental design that seeks to establish provable results. The absence of an explicit null hypothesis reinforces that the study is exploratory rather than confirmatory, meaning it is not designed to establish causality.

These issues collectively limit the work's scientific authority and the ability to form judgements about the effect of AI usage based on the results.

A More Accurate Framing of the Findings

Given the concerns—and criticisms—outlined here, the only defensible conclusion from this study is that

"Participants using AI during a writing task exhibited different neural activation patterns, recall performance, and writing characteristics compared with participants writing without AI."

Nothing more definitive follows. Nothing more can be claimed.

Whether AI harms cognition remains an open, nuanced, and highly context-dependent question. Thoughtful instructional design, assessment practices, and AI-usage frameworks matter far more than this study acknowledges.

Overall, while the study provides an interesting look at human–AI interaction, it does not provide clear evidence to support its central conclusions. It does raise questions worth further investigation, but should not be treated as definitive or policy-shaping evidence.

As with much early-stage research, the appropriate response is neither panic nor premature certainty, but curiosity and continued investigation.

Why the Study Is Attracting (Inappropriate) Attention

Given these concerns about the study design, results, and interpretation, why is this study garnering so much attention? It appears that its findings are being accepted without scrutiny by many (most?) who have reported on it or who have pointed to its findings when discussing AI usage by students and others. There may be several reasons for this.

Certainly, the MIT label confers prestige and suggests both credibility and authority, whether or not the study merits such acclaim. This may be a case in which prestige sometimes overwhelms the need for scrutiny and skepticism that is essential to scientific research. In a sense, those who report on this study may have the bias that if the study comes from MIT, it must be high quality with true conclusions. This is not a criticism of MIT or the authors, but a reminder that institutional prestige should never substitute for methodological scrutiny.

Reliance on prestige is especially risky when results have the potential to establish mindsets and influence policy.

Educators and others concerned or fearful about AI's role in learning may also be experiencing confirmation bias. This highly technical-looking paper from a respected institution appears to

validate existing fears, leading to confirmation even if the methodology does not support the claims.

Concluding Thought

The purpose of my critique is, in great part, an attempt to demonstrate that knowledge is not yet fixed, that any claims should not be accepted at face value, and that multiple perspectives are possible.

I offer this discussion not as a definitive, authoritative analysis but as one person's reflection and attempt to think critically about new information that may influence our access to opportunities and tools in an evolving world.

I welcome your feedback.

David Bowman, EdAINow
January 3, 2026