Effects of LLM use on Critical Thinking and High-Level Cognitive Functions

1st Steven Duval
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA
sduval8@gatech.edu

Abstract*—This literature review examines the long-term effects of LLM usage on critical thinking and its underlying cognitive components, executive function and metacognition, and how these effects vary across different LLM usage types. Through a careful review of recent academic literature, this paper identifies a consistent divide between two modes of LLM use that produce significantly different cognitive outcomes. Unstructured use, which constitutes the vast majority of current LLM interaction, is associated with significant gains in task-time performance but degraded critical thinking and higher-order cognition over time. Structured use, which scaffolds rather than replaces the user's reasoning process, shows the inverse pattern, reduced immediate efficiency paired with improved long-term metacognitive and critical thinking capacity. Three theoretical frameworks, namely cognitive offloading theory, cognitive load theory, and dual-process theory, help to explain these divergent outcomes through the suppression or preservation of effortful cognitive processing. The findings characterize an efficiency-proficiency tradeoff with implications for individuals, educators, and policymakers, and suggest that the long-term cognitive impact of LLMs depends not on whether they are used at all but on how they are operationalized.

***Keywords*—HCI, AI, LLM, Executive Function, Metacognition, Critical Thinking, Cognitive Offload Theory, Cognitive Load Theory, Dual Process Theory.**

I. Introduction

Large Language Model (LLM) use is pervasive, as of the beginning of 2025, over half of all American adults have at some point used LLMs [1]. These systems have demonstrated measurable impact in academia [2], consumer, organization, and government writing domains [3], and even in patterns of spoken language use among the general population [4]. This widespread influence amplifies the need to understand the long-term impacts, particularly on human cognition. Humans possess a unique ability to regulate one's own thinking and to think critically, capacities that quite possibly make us distinctively human [5]. These capacities are so important to our daily lives that they actually predict positive life outcomes better than intelligence [8]. Metacognition and executive function serve as fundamental components in the regulatory process underlying critical thinking [6] [7]. Therefore, understanding how different LLM usage types affect all three of these regulatory functions is paramount.

The different ways in which an LLM can be operationalized are not all created equal. This study explores these differences and distinguishing uses that foster the higher-order cognitive capacities from those that suppress them. This knowledge can equip us with the right conceptual frameworks in order to make informed decisions that impact us as human beings in the age of AI.

The profusion of accounts of the theoretical implications of the age of AI and LLMs read as though they are straight out of science-fiction novels [18][22][24][26][28][29][30]. Whether these arguments and projected models lean toward techno-optimism or straddle the line of Luddism, the message remains clear, it's imperative that we investigate the potential impacts these technologies have on humans in order to foster human flourishing and avoid suffering. As LLMs become increasingly embedded in everyday life, shaping how people communicate, learn, reason, and make decisions, understanding their effects on human cognition is no longer a peripheral concern but a central scientific responsibility. Additionally, the stakes of this question extend beyond the individual. LLMs are the first widely adopted tool capable of offloading not only memory but the higher-order reasoning processes. At the societal level, these offloading risks eroding the collective knowledge base that depends on individual effortful thinking [28]. Rigorous empirical investigation is essential to actively inform the development of systems that boost collective human flourishing.

This review adopts a cognitive science framing grounded in the computational-representational understanding of mind (CRUM) [9], which conceptualizes cognition in terms of knowledge representations such as beliefs, arguments, and knowledge and the computational processes performed over them, such as reasoning, evaluation, and explanation. Under this framework, critical thinking can be understood as the skillful execution of computation over mental representations [11], and metacognition and executive function serve as the regulatory mechanisms that enable that execution [12] [13].

The research question connects to this cognitive science framing in three fundamental ways:

1) *Critical thinking* as the manifestation of skillful computation over mental representations. *[11]*

2) *Distributed cognition and extended mind* as a framework for understanding how human and LLMs interact and effect each other [10] [17].

3) *Mechanistic theories* from cognitive science as the grounds for the effects to human long-term higher-order

cogntion, namely cognitive offloading theory [14], cognitive load theory [15], and dual-process theory [16].

II. Literature Review Design

This review was conducted through the careful examination of the budding academic research literature on LLMs and long-term impacts to human cognition. The research question was examined by synthesizing empirical findings across various study designs, including ex post facto surveys, classroom studies, and neuroimaging studies to identify patterns in how different LLM usage types affect critical thinking and its underlying cognitive components. The approach taken was iterative, starting with the broad topic of LLMs and critical thinking and progressively narrowing and refining the scope until arriving at a final research question both narrow enough to be meaningful yet broad enough to avoid esotericism.

Literature selection began with semantic search on Google Scholar and Scopus, supplemented by the literature mapping tools ResearchRabbit and Litmaps. These literature maps made it clear where the domain's foundational landmarks were located. A high signal selection tactic was taking note of commonly cited papers and authors in the sources' Related Works sections. This provided important sources that contributed foundational research in the domain. Sources were organized in Zotero by thematic groupings according to usage effects on higher-order cognition:

- Structured usage effects
- Unstructured usage effects
- Mechanisms
- Supplemental

Additionally, Notion pages were used to keep track of important findings and quotes for each paper. A higher-level page was used to track patterns, limitations, contradictions, and common conclusions.

Refinement followed a progression from abstract and conclusion screening for general relevance, to methods verification to confirm that the study's LLM usage was clear, to full-text review for extraction of key findings. Papers that were not cited or referenced by other papers in the domain were deprioritized, as a lack of engagement from the research community signaled limited influence in the field. Additionally, argument and prediction papers were removed from inclusion in the Results Section in order to keep a strict empirical grounding. A small collection of those theoretical papers were persevered and used as references to implications in the Discussion and Introduction sections.

There were three important focus shifts during the review process. First, in order to ground critical thinking in fundamental cognitive processes, metacognition and executive functions were added to the list of entities affected. This expanded the scope of the investigation to effects on all three. Second, most of the original potential sources looked at the task-time effects of using an LLM to do critical thinking work. The primary intention of this literature review is to investigate the effects on human's ability to critically reason in the long-run, particularly without the aid of the LLM tool. Therefore, there was a shift in focus from overall general effects to a distinction between short-term task-time effects and long-term post-task effects, with emphasis on the latter. Third, my original conceptualization of LLM usage as a single unit morphed over time into the understanding that the operationalization of the LLM is what matters, which resulted in the addition of usage types to the research question. From this addition, I had expectations of creating a sort of anthology of the various ways in which an LLM can be operationalized and each of their individual effects on critical thinking. Reading through the research, it became apparent that the important differentiation was not the specific granular usage type but rather whether the LLM was structured at all.

III. Results

An initial review of the literature on LLM effects on critical thinking reveals a seemingly highly contradictory environment. LLMs can somehow both foster and impair long-term higher-order functioning. Upon further analysis a distinctive pattern that hangs on a single dividing variable, the operationalization of the LLM, was discovered. Two general usage types emerge consistently across studies, unstructured use, which accepts the default "AI assistant" setup and structured use, which tailors responses to scaffold specific user effort. Each producing significantly different long-term cognitive outcomes.

A. Unstructured LLM usage and its long-term effects

LLM adoption rates have been one of the fastest in the history of digital technology [31]. Chatterji in collaboration with OpenAI researchers investigated ChatGPT usage data over 2 years. ChatGPT's weekly active users have grown from 1 million in 2023 to 800 million by late 2025. This data also established a window into the scale at which unstructured LLM interaction occurs today. Analyses of the conversation content shows that nearly 80% of interactions revolve around practical guidance, information retrieval, or writing tasks [31]. Suggesting the vast majority of ChatGPT conversations are exercises in the delegation of critical reasoning, metacognition, and executive functioning.

Any user of today's models, regardless of how much they use it, can tell you of the borderline-addicting productivity benefits of using a LLM and the task-time research evidence confirms these measurable gains. Not only do users with LLM access complete reasoning tasks significantly faster than those without, but qualitative assessments suggest higher output quality in complex ideation and data interpretation tasks [25] [32]. Notable quality exceptions do exist, such as originality, where evidence of the convergence of ideas and reduced creativity has been observed [33] [19] [25] [24].

These task-time gains do not appear to persist after the tool is removed. Across various study designs the evidence shows a consistent pattern of unstructured LLM use being associated with impairment of individual higher-order cognition over time. Ex post facto survey data reveals that higher self-reported LLM usage correlates with lower scores on memory retention, independent reasoning strategies, and critical thinking assessments [25]. Gerlich corroborates this at scale, finding a significant negative correlation between frequent AI/LLM tool usage and critical thinking abilities, with younger and more frequent users showing the sharpest declines [34]. Classroom studies reinforce the pattern across semester-long timescales and pedagogical domains. Li found that students who primarily used LLMs for academic work showed diminished ability to construct and evaluate arguments independently over the course of a semester [24], while Jošt et al. demonstrated that in programming course, students who relied more heavily on LLMs for critical thinking-intensive tasks such as code generation and debugging, performed significantly worse on an final assessment when the LLM was removed [23]. Stadler et al. demonstrated that while LLMs reduce mental load during learning tasks, this cognitive ease comes at the cost of processing depth. Students failed to develop the deep encoding needed for understanding [36]. Zhai similarly found that the reliance on LLMs, generated by efficiency gains and trust in quality outputs, erodes the user's capacity for independent critical assessment [32].

In the absence of longitudinal studies (see Limitations section for further details) neuro-activity monitoring offers the clearest evidence of higher-order cognition impairments of unstructured LLM use. Using EEG technology, Kosmyna et al. monitored brain activity and connectivity during writing tasks and found that LLM-assisted writers exhibited significantly less neural engagement compared to unassisted writers, with the greatest torpidity occurring in regions associated with executive function and metacognitive self-monitoring [19]. The study also revealed that LLM use fundamentally transformed the cognitive signature of the writing task itself. What was previously an act of critical composition and reasoning, writing became a task dominated by procedural integration and motor coordination [19].

B. Structured LLM usage and its long-term affects

Structured LLM use incorporates designs or processes that scaffold, rather than replace, the user's reasoning process. The most studied implementation is the Socratic chatbot in academic environments. This design is a model prompted or fine-tuned to respond with reflective questions rather than direct answers. When a conventional LLM-based writing tool was compared against a reflective-prompting LLM, the standard tool produced faster task completion but worse long-term skill development, while the reflective condition showed slower task performance but improved metacognitive engagement and independent reasoning over time [44]. These Socratic designs have been shown to consistently promote argument analysis, evidence evaluation, and deeper knowledge encoding [20] [45] [46] [47] [48].

These fostering effects emerge across cognitive components. A four-month study using an LLM-based metacognitive reflection tool produced measurable gains in students' ability to monitor and regulate their own thinking, with effects persisting beyond the study period [49]. Similarly, chatbot systems designed with explicit executive function scaffolding, which required users to plan, monitor, and evaluate responses, demonstrated improvements in working memory and cognitive flexibility [21].

Notably, even lightweight interventions show similar results. Lira et al. found that a coaching-aligned LLM, that prompted users to attempt tasks before receiving assistance, improved writing skill over time despite reducing short-term output efficiency [53]. These findings align with Tankelevitch et al.'s framing of generative AI as an inherently metacognitively demanding process when used properly. This process requires users to evaluate outputs, calibrate their knowledge, and monitor their own understanding [35]. The suggestion from these structured designs are that they foster critical thinking precisely because they force these demands onto the user rather than leaving it to the user's discretion.

However, adoption of structured LLM modes remains low. Darvishi et al. found that students consistently chose to rely on LLM outputs rather than learn from available AI-assisted scaffolds [37]. Similarly, Han discovered even when made explicitly aware of structured course-aligned LLM modes, students tend to opt for the default general-purpose LLM modes 65% of the time [38]. These results suggest that behavioral friction may limit the adoption of structured designs.

C. Mechanisms

Three theoretical frameworks work together, and overlap, to explain the divergent outcomes of unstructured and structured LLM use.

The most empirically supported is cognitive offloading theory. Risko and Gilbert define cognitive offloading as the use of physical action or external artifacts to reduce the cognitive demands placed on the internal system [14]. The consequences of offloading have been well-documented, internet search, digital photography, GPS navigation, and cockpit automation have all been shown to enhance task-time performance while causing atrophy to the capacities they replace [39] [40] [41] [42] [43]. LLMs extend this dynamic by enabling the offloading of higher-order reasoning itself. Gerlich demonstrated that cognitive offloading functions as the primary mediating factor between unstructured LLM use and long-term critical thinking impairment [34]. Gilbert's subsequent work frames the offloading decision as explicitly value-based, meaning users offload cognition when the perceived performance gain outweighs the cost effort, a perception that modern LLMs easily provide. [50]. Kosmyna et al.'s neuroimaging data confirms that offloading writing onto an LLM skipped the effortful cognitive processing required for deep encoding [19].

Cognitive load theory provides a complementary mechanistic account of effects [15]. LLMs greatly reduce extraneous cognitive load by automating summarization and information retrieval, which theoretically frees working memory for schema construction. However, unstructured usage also undermines the germane load that is needed to integrate new information into existing knowledge schemas. This results in surface-level understanding without permanent schema updates [36] [51]. Structured LLM designs, by contrast, preserve germane load through scaffolded prompting that keeps the user as the primary provider of cognitive effort while dynamically reducing extraneous load [52].

Dual-process theory offers a third, behavioral frame [16]. By surfacing seemingly coherent responses rapidly, LLMs systematically raise the threshold for the System 2 override, which is slow and deliberate processing, in favor of habitual use of System 1, which is fast, heuristic, low-effort processing. Over repeated interactions, this pattern may bolster the user's tendency for reactive processing rather than effortful cognition that's needed for critical analysis. Structured LLM interactions, which require the user to generate and evaluate their own reasoning, require System 2 engagement and instill the cognitive habits on which critical thinking depends.

IV. Disscussion

The findings in this review carry serious implications that extend far beyond the academic environment into the rest of the real world. The unstructured-structured distinction is not an ontological contribution, it conceptualizes a fault line along which LLM usage can either erode or foster the higher-order cognitive capabilities that constitute critical thinking. The behavioral evidence suggests that users overwhelmingly default to the side of that fault line associated with long-term erosion. This has significant and pressing implications to the future of education, policy, and LLM design.

A. Implications for Individuals and Society

At the individual level, the corpus of evidence paints a concerning vignette. The pattern of measurable gains in task-time performance at the expense of higher-order functioning gives fuel to Kim et al.'s behavioral predictions, in which sustained reliance on LLMs transitions from algorithm aversion to algorithm dependence, progressively diminishing the user's self-efficacy and willingness to engage in effortful reasoning [26]. What makes this progression especially concerning is its self-reinforcing properties. As noted in the results section, Zhai found that the overreliance of LLMs erodes the user's capacity for independent critical assessment [32], while also noting that this overreliance arises when individuals struggle to critically asses the reliability of AI. This suggests a negative feedback loop in which unstructured LLM use diminishes the critical faculties needed to evaluate one's own LLM use, driving further dependence. García-Barrios formalizes this self-reinforcement dynamic through a conceptual model of cognitive offloading in human-AI collaboration, identifying the conditions where intelligent systems transition from epistemic partners to cognitive crutches [27]. This model suggests that this transition is not sudden but gradual. Users will pass through stages of increasing dependence where each stage of offloading makes the next more likely, precisely because the metacognitive resources needed to resist further offloading are themselves being eroded.

Beyond the individual there are also societal implications. Acemoglu, Kong, and Ozdaglar formalize the collective impact of this problem through a dynamic model of learning and decision-making in which individual incentives and effort produce knowledge that contributes to a shared stockpile of human knowledge [28]. Their results demonstrates that when AI provides sufficiently accurate information, it replaces the human effort that sustains collective knowledge, and the system can tip into what they term a knowledge-collapse steady state, which is a long-run equilibrium state in which society's stockpile of general knowledge completely vanishes despite the access to high-quality personalized AI advice. This review could provide concrete mechanisms by which the cognitive effort replacement dynamics in their model would operate. Taken together, the theoretical and empirical evidence suggests that unstructured LLM use poses risks not only to individual proficiency but to the collective society.

B. Significance for Cognitive Science and Real-World Application

The contents of this review matter and have implications on cognitive science for multiple reasons. First, the research provides further empirical support for cognitive offloading theory by demonstrating its operation in a new and rapidly evolving technological domain. More specifically, LLMs represent a new case in which not only mental representations like memories can be offloaded to an external artifact, but the reasoning computations performed over those representations as well. Under the CRUM framework adopted in this review, this amounts to the offloading of both the representations and the computations of cognition.

This externalization naturally leads to implications in distributed cognition [10] and the extended mind theory [17], raising the question of whether LLMs are best understood as tools that scaffold cognition or as tools that extend the boundary of it. The evidence from this review suggests that the answer depends on operationalization. The structured use of LLMs may genuinely bolster cognitive capacity in the way the extended mind thesis envisions, while unstructured use may be better thought of as a replacement that progressively atrophies our own cognitive capacity. If we continue on the usage trajectory we are on one could argute that it makes sense to incorporate LLMs as a fundamental component of human cognition.

In real-world contexts, the recurring pattern in the research of unstructured use maximizing immediate task performance while structured use supports long-term capability can be conceptualized as an efficiency-proficiency tradeoff. Recognizing this tradeoff provides a decision framework for various domains in which LLMs are used. In education, it's possible the value of an LLM interaction depends on whether the goal is task completion or skill proficiency. Organization of academic resources might be a suitable task for efficiency while the fundamental knowledge encoding of concepts is better suited for proficiency. In professional contexts such as legal analysis, medical reasoning, and strategic decision-making, practitioners must assess whether a given task demands independent expertise or whether the delegation to an LLM has acceptable cognitive cost. A financial analyst who offloads modeling to an LLM may produce faster deliverables and which might translate to capital gains but fail to develop the judgment that comes from wrestling with the analysis.

C. Implications for Education, Policy, and Design

This efficiency-proficiency tradeoff has direct implications for how LLMs are integrated going forward. In education, the evidence suggests that unrestricted LLM access in learning environments risks undermining the skill development that education is meant to provide. The structured LLM usages reviewed demonstrate that it is possible to design or operationalize LLMs to enhance learning outcomes. So far, the adoption evidence from Darvishi et al. and Han indicates that students will not gravitate toward these modes voluntarily [37] [38]. This places a responsibility on both students and institutions to determine how LLMs are used and in what context.

For policymakers, the societal risks formalized by Acemoglu et al.'s knowledge-collapse model suggest that the question of LLM regulation should be considered. If unstructured LLM use at scale erodes the collective capacity for independent reasoning, then policies governing LLMs must account for externalities.

The most responsibility may lie with the LLM designers. The default configuration of current LLMs is the setup most associated with cognitive offloading and long-term impairment. Structured designs that scaffold rather than replace reasoning exist and work, but they require deliberate implementation on the user's part. The gap between what is technically possible and what users actually adopt suggests that effective structured LLM design cannot rely on user initiative alone but must be embedded into the default interaction.

V. Conclusion

I began this investigation with the assumption that LLM usage was relatively uniform whose effects could be assessed as a whole. What I learned is that this framing completely hides the most important detail. When it comes to cognition, the structured-unstructured LMM usage distinction is not a minor methodological detail, it is the central factor of whether LLM use fosters or impairs the cognitive abilities that matter most. More importantly, I came away with the understanding that LLMs are not inherently offloading tools. They can be deliberately designed or used to scaffold reasoning rather than replace it.

The central finding of this review is that the long-term cognitive effects of LLM usage on critical thinking and its underlying components hinge on how the model is operationalized. Though longitudinal studies are greatly in need, the available evidence points to a divide between two usage modes that produce markedly different cognitive outcomes.

Unstructured LLM use, the default mode adopted by most users, significantly increases task-time performance and efficiency in critical thinking, metacognitive, and executive function tasks, but significantly decreases the user's ability to perform those tasks on their own over time. Cognitive offloading functions as the primary mediating factor, with support from cognitive load theory and dual-process theory reinforcing the pattern. Structured LLM use produces the inverse, reduced immediate task efficiency paired with stronger encoding of the reasoning processes that constitute critical thinking.

This distinction reframes the broader conversation around LLMs and cognition. The question is not whether LLMs themselves affect critical thinking but how we choose to use them in specific contexts. The efficiency-proficiency tradeoff identified in this review provides a decision framework for navigating that choice. It implies that every LLM interaction involves a decision between optimizing for immediate output and investing in long-term cognitive capability. How we all as individuals, students, and professionals negotiate that tradeoff will have serious implications on determining whether LLMs serve as tools that support human reasoning or systems that replace it.

VI. Limitations

The most significant limitation of this review is the serious lack of longitudinal studies across both structured and unstructured usage domains. While LLMs have been publicly accessible since late 2022, their reasoning capabilities have only recently developed to a point where users rely on them for higher-order cognitive work. The initial breakthrough, OpenAI's o1 reasoning model, which introduced the concept of chain-of-thought reasoning, reached general availability in late 2024 [54]. This timeline has left researchers an insufficient amount of time to conduct rigorous longitudinal studies on the long-term cognitive effects of reasoning LLMs. As a result, the long-term conclusions drawn in this review rest on ex post facto survey designs, semester-length classroom studies, and neuroimaging snapshots.

The empirical literature on structured LLM use is concentrated in educational settings, making it hard to extrapolate the fostering effects to different domains. Additionally, there is no known comparative studies that systematically evaluate multiple structured LLM designs against one another. The present review treats structured use as a general category, but the specific scaffolding approach likely matters, and no existing research maps these variations against one another within a unified framework.

Future research should prioritize these directions. First and foremost longitudinal studies that track everyday LLM users over years would provide the empirical foundation that the current literature lacks. Comparative studies evaluating different structured LLM designs would clarify which scaffolding approaches produce the strongest outcomes and in which domains. Neuroimaging methodologies, which provided some of the most compelling evidence in this review, should continue to be integrated as the technology improves. Finally, behavioral studies examining how people naturally operationalize LLMs outside of controlled settings would ground the structured-unstructured distinction in real-world usage patterns.

References

[1] L. Rainie, "Close encounters of the AI kind: The increasingly human-like way people are engaging with language models," Imagining the Digital Future Center, Elon University, Mar. 2025. [Online]. Available: https://imaginingthedigitalfuture.org/reports-and-publications/close-encounters-of-the-ai-kind/close-encounters-of-the-ai-kind-main-report/

[2] W. Liang *et al.*, "Mapping the increasing use of llms in scientific papers," arXiv:2404.01268, 2024.

[3] W. Liang, Y. Zhang, M. Codreanu, J. Wang, H. Cao, and J. Zou, "The widespread adoption of large language model-assisted writing across society," arXiv:2502.09747, 2025.

[4] H. Yakura, E. Lopez-Lopez, L. Brinkmann, I. Serna, P. Gupta, and I. Rahwan, "Empirical evidence of large language model's influence on human spoken communication," arXiv:2409.01754, 2024.

[5] M. C. Corballis, "The uniqueness of human recursive thinking: The ability to think about thinking may be the critical attribute that distinguishes us from all other species," *American Scientist*, vol. 95, no. 3, pp. 240-248, 2007.

[6] C. Magno, "The role of metacognitive skills in developing critical thinking," *Metacognition and Learning*, vol. 5, no. 2, pp. 137-156, 2010.

[7] S. Li, X. Ren, K. Schweizer, T. M. Brinthaupt, and T. Wang, "Executive functions as predictors of critical thinking: Behavioral and neural evidence," *Learning and Instruction*, vol. 71, p. 101376, 2021.

[8] H. A. Butler, C. Pentoney, and M. P. Bong, "Predicting real-world outcomes: Critical thinking ability is a better predictor of life decisions than intelligence," *Thinking Skills and Creativity*, vol. 25, pp. 38-46, 2017

[9] P. Thagard, "Mind: Introduction to cognitive science," MIT Press, 2005.

[10] E. Hutchins and T. Klausen, "Distributed cognition in an airline cockpit," in *Cognition and Communication at Work*, p. 15, 1996.

[11] Foundation for Critical Thinking, "Defining Critical Thinking." [Online]. Available:https://www.criticalthinking.org/pages/defining-critical-thinking/766. [Accessed: Feb. 8, 2026].

[12] A. P. Shimamura, "Toward a cognitive neuroscience of metacognition," *Consciousness and Cognition*, vol. 9, no. 2, pp. 313-323, 2000.

[13] M. T. Banich, "Executive function: The search for an integrated account," *Current Directions in Psychological Science*, vol. 18, no. 2, pp. 89-94, 2009.

[14] E. F. Risko and S. J. Gilbert, "Cognitive offloading," *Trends in Cognitive Sciences*, vol. 20, no. 9, pp. 676-688, 2016.

[15] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, Apr. 1988, doi: 10.1207/s15516709cog1202_4

[16] D. Kahneman and S. Frederick, "A model of heuristic judgment," in *The Cambridge Handbook of Thinking and Reasoning*, K. Holyoak and B. Morrison, Eds. Cambridge, U.K.: Cambridge University Press, 2005, pp. 267–293

[17] A. Clark and D. J. Chalmers, "The extended mind," *Analysis*, vol. 58, no. 1, pp. 7–19, Jan. 1998, doi: 10.1093/analys/58.1.7.

[18] S. Grinschgl and A. C. Neubauer, "Supporting cognition with modern technology: Distributed cognition today and in an AI-enhanced future," Frontiers in Artificial Intelligence, vol. 5, Art. no. 908261, 2022.

[19] N. Kosmyna et al., "Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task," arXiv preprint, arXiv:2506.08872, 2025.

[20] R. A. Fabio, A. Plebe, and R. Suriano, "AI-based chatbot interactions and critical thinking skills: An exploratory study," Current Psychology, pp. 1–14, 2024

[21] P. Pergantis, V. Bamicha, C. Skianis, and A. Drigas, "AI chatbots and cognitive control: Enhancing executive functions through chatbot interactions: A systematic review," Brain Sciences, vol. 15, no. 1, p. 47, 2025

[22] U. León-Domínguez, "Potential cognitive risks of generative transformer-based AI chatbots on higher order executive functions," Neuropsychology, vol. 38, no. 4, p. 293, 2024

[23] G. Jošt, V. Taneski, and S. Karakatič, "The impact of large language models on programming education and student learning outcomes," Applied Sciences, vol. 14, no. 10, p. 4115, 2024

[24] B. Li, "Study on the long-term cognitive effects of excessive LLM use among students and general users," Journal of Contemporary Educational Research, vol. 9, no. 10, 2025, doi: 10.26689/jcer.v9i10.12112

[25] S. Jamshaid, "Cognitive effects of large language models: An empirical study of memory retention, reasoning strategies, and critical thinking among students and professionals at the University of Gujrat, Pakistan," ACADEMIA International Journal for Social Sciences, vol. 4, no. 3, 2025, doi: 10.63056/acad.004.03.0893

[26] T. W. Kim, U. Usman, A. Garvey, and A. Duhachek, "From algorithm aversion to AI dependence: Deskilling, upskilling, and emerging addictions in the GenAI age," Consumer Psychology Review, 2026, doi: 10.1002/arcp.70008

[27] D. A. García-Barrios, "Epistemic partner or cognitive crutch: A conceptual model of cognitive offloading in human–artificial intelligence collaboration," *IEEE Internet Computing*, vol. 29, no. 5, pp. 7–16, Sep./Oct. 2025, doi: 10.1109/MIC.2025.3626694.

[28] D. Acemoglu, D. Kong, and A. Ozdaglar, "AI, human cognition and knowledge collapse," National Bureau of Economic Research, Working Paper 34910, Feb. 2026. [Online]. Available: https://economics.mit.edu/sites/default/files/2026-02/AI,%20Human%20Cognition%20and%20Knowledge%20Collapse%2002-20-26.pdf

[29] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in *Proc. 2021 ACM Conf. Fairness, Accountability, and Transparency (FAccT '21)*, New York, NY, USA: ACM, Mar. 2021, pp. 610–623, doi: 10.1145/3442188.3445922.

[30] S. Wachter, B. Mittelstadt, and C. Russell, "Do large language models have a legal duty to tell the truth?" *R. Soc. Open Sci.*, vol. 11, no. 8, p. 240197, Aug. 2024, doi: 10.1098/rsos.240197

[31] A. Chatterji, T. Cunningham, D. J. Deming, Z. Hitzig, C. Ong, C. Y. Shan, and K. Wadman, "How People Use ChatGPT," National Bureau of Economic Research, Working Paper 34255, Sep. 2025, doi: 10.3386/w34255

[32] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review," *Smart Learning Environments*, vol. 11, no. 1, p. 28, 2024, doi: 10.1186/s40561-024-00316-7

[33] B. Jose, J. Cherian, A. M. Verghis, S. M. Varghise, M. S, and S. Joseph, "The cognitive paradox of AI in education: between enhancement and erosion," *Frontiers in Psychology*, vol. 16, 2025, doi: 10.3389/fpsyg.2025.1550621

[34] M. Gerlich, "AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking," *Societies*, vol. 15, no. 1, p. 6, 2025, doi: 10.3390/soc15010006.

[35] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, and S. Rintel, "The Metacognitive Demands and Opportunities of Generative AI," in *Proc. 2024 CHI Conf. Human Factors Comput. Syst.*, New York, NY, USA: Association for Computing Machinery, 2024, pp. 1–24, doi: 10.1145/3613904.3642902.

[36] M. Stadler, M. Bannert, and M. Sailer, "Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry," *Computers in Human Behavior*, vol. 160, p. 108386, 2024, doi: 10.1016/j.chb.2024.108386

[37] A. Darvishi, H. Khosravi, S. Sadiq, D. Gašević, and G. Siemens, "Impact of AI assistance on student agency," *Computers & Education*, vol. 210, p. 104967, 2024, doi: 10.1016/j.compedu.2023.104967.

[38] S. Han, "Learner autonomy and mode selection in dual-mode chatbot: Students' interaction patterns and learning outcomes in an online course," *Computers & Education*, vol. 240, p. 105464, 2026, doi: 10.1016/j.compedu.2025.105464.

[39] B. Sparrow, J. Liu, and D. M. Wegner, "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips," *Science*, vol. 333, no. 6043, pp. 776–778, 2011, doi: 10.1126/science.1207745.

[40] M. Fisher, M. K. Goddu, and F. C. Keil, "Searching for explanations: How the Internet inflates estimates of internal knowledge," *Journal of Experimental Psychology: General*, vol. 144, no. 3, pp. 674–687, 2015, doi: 10.1037/xge0000070.

[41] L. A. Henkel, "Point-and-Shoot Memories: The Influence of Taking Photos on Memory for a Museum Tour," *Psychological Science*, vol. 25, no. 2, pp. 396–402, 2014, doi: 10.1177/0956797613504438.

[42] A. L. Gardony, T. T. Brunyé, C. R. Mahoney, and H. A. Taylor, "How Navigational Aids Impair Spatial Memory: Evidence for Divided Attention," *Spatial Cognition & Computation*, vol. 13, no. 4, pp. 319–350, 2013, doi: 10.1080/13875868.2013.792821.

[43] S. M. Casner, R. W. Geven, M. P. Recker, and J. W. Schooler, "The Retention of Manual Flying Skills in the Automated Cockpit," *Human Factors*, vol. 56, no. 8, pp. 1506–1516, 2014, doi: 10.1177/0018720814535628

[44] C. Liu, J. Hou, Y.-F. Tu, Y. Wang, and G.-J. Hwang, "Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments," *Interactive Learning Environments*, vol. 31, no. 9, pp. 5614–5632, 2023, doi: 10.1080/10494820.2021.2012812

[45] R. Suriano, A. Plebe, A. Acciai, and R. A. Fabio, "Student interaction with ChatGPT can promote complex critical thinking skills," *Learning and Instruction*, vol. 95, p. 102011, 2025, doi: 10.1016/j.learninstruc.2024.102011.

[46] L. Xi, Y. Zhang, and Q. Wang, "Investigating the effects of an LLM-based Socratic conversational agent on students' academic performance and reflective thinking in higher education," *Computers & Education*, vol. 241, p. 105494, 2026, doi: 10.1016/j.compedu.2025.105494

[47] X. Xie, X. Yang, and R. Cui, "A Large Language Model-Based System for Socratic Inquiry: Fostering Deep Learning and Memory Consolidation," in *Proc. 2025 14th Int. Conf. Educational and Information Technology (ICEIT)*, 2025, pp. 52–57, doi: 10.1109/ICEIT64364.2025.10975940

[48] H. Kumar et al., "Supporting Self-Reflection at Scale with Large Language Models: Insights from Randomized Field Experiments in Classrooms," in *Proc. Eleventh ACM Conf. Learning @ Scale*, New York, NY, USA: Association for Computing Machinery, 2024, pp. 86–97, doi: 10.1145/3657604.3662042

[49] F. Fan, S. Wang, M. D. Mai, X. Nie, and L. T. Yang, "Enhancing Students' Metacognition With Innovative IA-Based Metacognitive Reflective Learning Tool," *IEEE Transactions on Learning Technologies*, vol. 18, pp. 699–715, 2025, doi: 10.1109/TLT.2025.3580536.

[50] S. J. Gilbert, "Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory," *Cognition*, vol. 247, p. 105783, 2024, doi: 10.1016/j.cognition.2024.105783.

[51] S. Grinschgl, F. Papenmeier, and H. S. Meyerhoff, "Consequences of cognitive offloading: Boosting performance but diminishing memory," *Quarterly Journal of Experimental Psychology*, vol. 74, no. 9, pp. 1477–1496, 2021, doi: 10.1177/17470218211008060.

[52] E. Gkintoni, H. Antonopoulou, A. Sortwell, and C. Halkiopoulos, "Challenging Cognitive Load Theory: The Role of Educational Neuroscience and Artificial Intelligence in Redefining Learning Efficacy," *Brain Sciences*, vol. 15, no. 2, p. 203, 2025, doi: 10.3390/brainsci15020203.

[53] B. Lira, T. Rogers, D. G. Goldstein, L. Ungar, and A. L. Duckworth, "Coach not crutch: Evidence that AI can improve writing skill despite reducing effort," arXiv, preprint arXiv:2502.02880, 2026, doi: 10.48550/arXiv.2502.02880.

[54] OpenAI, "Learning to Reason with LLMs," Sep. 12, 2024. [Online]. Available: https://openai.com/index/learning-to-reason-with-llms/