

Thinking with AI or Thinking Less? A Narrative Review of Generative AI Dependence, Cognitive Offloading, and Human Cognitive Functioning

Deva Adithiya L.M.D.*; Samarawickrema N.S.; Peiris J.A.H.L.; Weerasinghe B.R.D.S
Department of Psychology, Faculty of Behavioural Sciences, KIU, Sri Lanka

Dissanayake C.T.N.
Department of Biomedical Science, Faculty of Health Sciences, KIU, Sri Lanka

*Corresponding author

Abstract

The new generation of digital assistance programs, including generative artificial intelligence (GenAI), is increasingly shifting from information retrieval to the generation of explanations, arguments, code, abstracts, and solutions. The extent to which GenAI support of human cognition can extend human capacity versus supplant critical cognitive processes underlying knowledge and skill formation has been a longstanding psychological question. This narrative review focuses on foundational research on cognitive offloading and on peer-reviewed studies of GenAI published through September 05, 2026. A structured narrative search for experimental, meta-analytic, and mixed-methods studies on memory, learning, critical thinking, creativity, perceived effort, trust, and dependence on GenAI was complemented by a search of cross-sectional and qualitative studies that provided early characterization of emerging constructs of dependence on GenAI. Evidence does not support a simple positive or negative effect of GenAI on human cognition. GenAI-supported offloading of cognitive functions such as memory retrieval, processing, and learning can reduce the required cognitive capacity and support high performance across a variety of tasks, including memory-based and GenAI-supported academic tasks. However, the same cognitive offloading can, in some cases, reduce the quality of subsequent encoding, and recent studies on GenAI show a consistent performance-learning dissociation. While support that enables high performance when support is available can carry over to later unsupported performance, the gain is typically small, nonexistent, or even negative. Studies of support provided under instructional control indicate that retaining learner-generated nodes or restricting direct-answer substitutes for generators can support later independent performance better than unlimited support. It remains to be determined what the independent contribution of explanation, verification, revision, and fading support is. Nevertheless, findings from emerging dependence research correlate with indicators of academic stress, performance expectations, anxiety, and self-efficacy. However, currently available evidence does not support claims of irreversible cognitive decline. A key finding of this review is that whether GenAI is used at all is less critical than which functions are offloaded, which are maintained, and the extent to which users remain epistemically responsible for the resulting output. Implications are developed for education, knowledge work, tool design, and longitudinal research.

Keywords:

Generative Artificial Intelligence; Cognitive Offloading; AI Dependence; Critical Thinking; Memory; Metacognition; Learning Transfer; Human-AI Collaboration.

1. Introduction

Generative artificial intelligence has altered the cognitive ecology of everyday work. Earlier digital tools mainly stored, searched, calculated, or reminded. Contemporary large language model interfaces can instead draft an argument, propose a diagnosis for a technical fault, generate a worked solution, summarize a literature set, rewrite prose, suggest code, or answer a question in conversational form. The user is therefore no longer offloading only storage and retrieval. Depending on the task, the system can take over portions of idea generation, organization, inference, explanation, evaluation, and decision preparation. This wider functional range makes a familiar psychological trade-off more consequential: external support may release scarce cognitive resources, yet it may also remove the very mental activity that produces learning.

The starting point is cognitive offloading, commonly defined as using external action or environmental resources to reduce the information-processing requirements of a task (Risko & Gilbert, 2016). Offloading is not intrinsically maladaptive. People write lists, set reminders, save files, use calculators, and consult reference sources because internal attention and memory are limited. Laboratory work shows that strategic reminders can improve prospective memory, that people adjust offloading according to task difficulty and their confidence in internal memory, and that externalizing one set of information can free resources for learning other material (Gilbert, 2015; Gilbert et al., 2020; Storm & Stone, 2015). A recent meta-analysis likewise found that cognitive offloading generally improves performance on memory-based tasks and can reduce individual differences in task success (Burnett & Richmond, 2026). Treating all offloading as cognitive deterioration would therefore conflict with a substantial empirical literature.

The difficulty lies in separating successful task completion from durable human capability. The distinction between performance and learning is well established: behavior observed during practice can improve without a corresponding durable change in knowledge or skill, while effortful conditions that depress immediate performance can sometimes produce stronger retention and transfer (Soderstrom & Bjork, 2015). This distinction is especially important for GenAI. A student who obtains a correct solution, an employee who produces a polished report, or a programmer who repairs code with an AI assistant may have achieved the task goal. Those outcomes do not, by themselves, show what the person can later recall, explain, detect, transfer, or reproduce without assistance.

Pre-GenAI evidence already anticipated this tension. When people expected information to remain externally available, they remembered less of the information itself and more about where it could be found (Sparrow et al., 2011). Internet search could also inflate judgments of internal knowledge, creating a gap between access to information and what a person actually knew (Fisher et al., 2015). Experimental work on offloading later showed that frequent externalization can improve immediate task performance while reducing subsequent memory for the offloaded material, including when participants know that memory will later be tested (Grinschgl et al., 2021). Expecting an external store can reduce study time and alter encoding strategies, and offloading has been associated with lower true recall and higher false recall when the store is unexpectedly unavailable (Kelly & Risko, 2022; Lu et al., 2020). These findings do not prove that GenAI has the same effects. They identify mechanisms that become plausible when a generative system supplies not only information, but also intermediate cognitive work.

The GenAI literature is now sufficiently developed to reject equally simplistic narratives at both extremes. Meta-analytic evidence indicates that ChatGPT and related systems can improve academic performance, affective-motivational outcomes, and some measures of higher-order thinking while reducing mental effort (Deng et al., 2025). A separate meta-analysis of university studies reported positive pooled effects on academic achievement and higher-order thinking, although effects on metacognition were not statistically significant and substantial heterogeneity remained across contexts (Chen & Cheung, 2025). At the same time, controlled studies have begun to show that unrestricted AI assistance can produce weaker unaided learning or retention than conditions that preserve self-generated reasoning (Barcaui, 2025; Bastani et al., 2025; Wong & Qiu, 2026).

The central question, then, is not whether humans should think with tools. They always have. The more precise question is which cognitive operations can be delegated with little cost, which operations need to remain internal for durable learning and autonomy, and under what conditions AI support changes from

scaffold to substitute. This review addresses four aims: (a) to connect GenAI research with the broader cognitive-offloading literature; (b) to distinguish immediate assisted performance from durable human learning and transfer; (c) to evaluate evidence concerning memory, critical thinking, creativity, metacognition, and emerging AI dependence; and (d) to propose an empirically constrained framework for identifying when GenAI use is more likely to scaffold cognition or replace it.

2. Review Methodology

This article uses a structured narrative review design oriented toward explanatory synthesis rather than exhaustive estimation of a single pooled effect. Reporting and internal quality checks were informed by the Scale for the Assessment of Narrative Review Articles (SANRA), which emphasizes justification of review importance, statement of aims, description of the literature search, appropriate referencing, scientific reasoning, and presentation of relevant endpoint data (Baethge et al., 2019). The review question was intentionally broad: how does GenAI-related cognitive offloading affect immediate performance, durable learning, memory, critical thinking, creativity, metacognitive calibration, and emerging dependence?

Searches were conducted from September 01 to September 05, 2026 across PubMed, ERIC, the ACM Digital Library, and publisher-hosted scholarly interfaces, supplemented by backward and forward citation checking from major reviews and high-information empirical papers. Search strings combined ("generative AI" OR ChatGPT OR "large language model") with cognitive offloading, dependence, overreliance, critical thinking, memory, retention, retrieval, cognitive load, mental effort, metacognition, creativity, learning, transfer, trust, and "appropriate reliance." Foundational searches combined cognitive offloading, external memory, retrieval practice, self-explanation, generative learning, automation bias, and trust calibration. The final search was completed on September 05, 2026.

Eligible sources were English-language peer-reviewed empirical studies, meta-analyses, systematic reviews, or foundational theoretical and experimental papers directly bearing on the mechanisms under review. GenAI-specific evidence published from 2023 through September 05, 2026 was prioritized; older evidence was retained when it established mechanisms such as external memory, retrieval practice, self-explanation, cognitive offloading, or automation reliance. Studies were excluded from substantive conclusions when they were opinion pieces, non-peer-reviewed reports, purely technical model evaluations without human cognitive outcomes, or neurobiological claims unsupported by direct data. Cross-sectional and qualitative studies were retained only for questions they could validly address, particularly dependence constructs, perceived effort, and user experience.

Evidence was weighted by design strength and fit to claim. Meta-analyses, randomized or quasi-randomized interventions, preregistered studies, and behavioral experiments received greatest weight for causal or near-causal statements; self-report and cross-sectional evidence was treated as associational. The final synthesis drew on 31 substantive peer-reviewed sources plus the SANRA methodological source. Studies were organized around immediate performance, memory and retention, critical thinking, metacognition and trust calibration, creativity and transfer, and dependence or overreliance, then interpreted through the cross-cutting distinction between substitution and scaffolding. Because the purpose was integrative narrative synthesis rather than exhaustive systematic effect estimation, no PRISMA flow diagram or pooled effect calculation was produced; search and eligibility procedures are reported to make the selection logic auditable.

3. Results: Narrative Synthesis

Across the literature, three findings recur. First, offloading is often adaptive for the immediate task. Second, gains observed while an external aid is available cannot be assumed to represent durable learning. Third, the cognitive effect of GenAI depends strongly on interaction design and user behavior. The same system can serve as a tutor, editor, search interface, generator, verifier, or answer engine, and those roles impose different demands on the user. Table 1 summarizes representative empirical studies that anchor the synthesis.

Table 1 Representative evidence informing the narrative synthesis

Study	Design / sample	Domain	Principal finding	Interpretive limitation
Sparrow et al. (2011)	4 experiments	External information availability and memory	Expected future access reduced recall for the information itself while improving memory for where it could be found.	Pre-GenAI external memory paradigm; does not test GenAI.
Grinschgl et al. (2021)	3 experiments, N = 172 in each final sample (total N = 516)	Cognitive offloading and later memory	Offloading improved immediate performance and generally reduced later memory. In Experiment 3, participants forced to offload but aware of a later memory test largely compensated for the memory cost.	Laboratory task; memory costs were conditional rather than inevitable, and generalization to complex knowledge work requires caution.
Deng et al. (2025)	Meta-analysis, 69 experimental articles	Student learning with ChatGPT	Overall positive effects on academic performance, affective-motivational outcomes, and higher-order thinking propensities; mental effort decreased.	Heterogeneous interventions and outcome measures; authors note methodological weaknesses in primary studies.
Chen & Cheung (2025)	Meta-analysis, 57 studies, 97 estimates	University learning outcomes	Positive pooled effects for language skills, academic achievement, affective-motivational outcomes, and higher-order thinking; metacognition effect was not significant.	Substantial heterogeneity; effect sizes varied by learners, measures, setting, and intervention features.
Bastani et al. (2025)	Field experiment, nearly 1,000 high-school students	Mathematics practice and subsequent unaided performance	AI assistance raised performance during practice. Unrestricted assistance produced worse later unaided performance than control; guardrailed tutoring largely mitigated the harm.	Specific mathematics context and tutor implementation.
Barcaui (2025)	Randomized controlled trial, n = 120	45-day knowledge retention	ChatGPT-assisted learners scored lower on the delayed retention test than the traditional-study group.	Single learning domain; delayed-test analysis involved fewer observations than initially randomized.
Lee et al. (2025)	Survey, 319 knowledge workers, 936 use cases	Critical-thinking effort in work	Higher confidence in GenAI was associated with less reported critical-thinking effort; work shifted toward verification, integration, and stewardship.	Self-report study; cannot establish objective cognitive decline or causality.

Study	Design / sample	Domain	Principal finding	Interpretive limitation
Essel et al. (2024)	Pretest-posttest control study, n = 125	Critical, creative, and reflective thinking	ChatGPT-supported instruction was associated with higher post-intervention scores on measured cognitive-skill scales.	Outcomes were largely scale-based and context-specific, not delayed unaided transfer tests.
Jain et al. (2026)	Comparative behavioral study; 124 initially considered, final N = 120 after outlier exclusion	Fallacy detection, mental effort, reasoning behavior	AI-assisted participants were more accurate and reported lower mental effort; qualitative evidence distinguished strategic engagement from complete delegation.	Short-term task; no long-term retention or transfer assessment.
Wong & Qiu (2026)	Intervention study, N = 196	Creative performance and later independent creativity	Free AI use improved assisted creativity but did not yield the strongest later independent performance. A "think first, ChatGPT later" condition produced better later independent originality.	Creative problem-solving context; transfer measured within the same broad domain.
Zhang et al. (2024)	Cross-sectional survey, n = 300	AI dependency in university students	Academic stress and performance expectations were implicated in AI-dependency pathways; participants also reported perceived negative consequences.	Cross-sectional and self-report; direction of causality cannot be established.
Wang et al. (2026)	Cross-sectional student survey	AI dependence, stress, burnout, anxiety	AI dependence was statistically associated with academic stress, burnout, anxiety, and lower self-efficacy within the tested model.	Cross-sectional mediation does not demonstrate temporal or causal mediation.

Note. Findings are interpreted according to design strength; cross-sectional and self-report studies are not treated as evidence of causal cognitive decline.

3.1 Cognitive offloading is an adaptive strategy, not a deficit by definition

Cognitive offloading emerges because internal cognitive resources are finite. The offloading literature therefore begins from a resource-management problem rather than a pathology model. In prospective-memory tasks, participants strategically create reminders and tend to offload more when memory demands increase or confidence in unaided performance falls (Gilbert, 2015; Gilbert et al., 2020). This behavior is often rational. If the external cue is reliable, it raises the probability that an intention will be completed. Similarly, saving previously studied information can improve learning of subsequently presented information by reducing interference, an effect Storm and Stone (2015) described as saving-enhanced memory.

Meta-analytic evidence reinforces the adaptive account. Burnett and Richmond (2026) found that offloading improves performance on memory-based tasks, with benefits particularly evident when offloading is required rather than merely optional. They also reported reduced interindividual variability in performance, suggesting that external supports can partially compensate for differences in internal memory resources. This point matters for GenAI debates because lower mental effort is not itself evidence of cognitive harm. If a tool removes low-value memory demands, routine formatting, or repetitive search work, the released capacity can be redirected toward interpretation, planning, or higher-order judgment. The empirical question is whether that redirection actually occurs.

Offloading becomes cognitively consequential when external support changes how information is encoded or how often a mental operation is practiced. Sparrow et al. (2011) showed that expected access to stored information alters what people remember: less of the content and more of its location. Kelly and Risko (2022) subsequently found that anticipated external-store availability changes study effort and strategy, providing direct support for an encoding-effort mechanism. Across three experiments, Grinschgl et al. (2021) found that greater offloading improved immediate task performance but generally reduced later memory for the offloaded information. Crucially, Experiment 3 showed that participants forced to offload could largely compensate for the memory cost when they knew memory would later be tested. This boundary condition indicates that externalization does not mechanically impair memory formation when an explicit learning goal preserves internal processing. Lu et al. (2020) further showed that offloading could reduce true recall and increase false recall when the external store was unavailable. Together, these studies establish a principle rather than a prohibition: externalization changes the allocation of cognitive work, and its later consequences depend partly on what the user continues to process internally.

3.2 What generative AI changes: from external storage to external generation

Traditional external memory systems preserve information the user or another source has already produced. GenAI can instead synthesize and transform information, propose novel content, and supply intermediate steps. This changes the unit of offloading. A user can offload not only remembering a fact, but also producing examples, structuring an argument, generating hypotheses, interpreting a passage, drafting a conclusion, or deciding what should be considered next. In workplace data, Lee et al. (2025) found that GenAI shifted users' reported critical-thinking activity away from information gathering and task execution toward verification, response integration, and task stewardship. That is not simply less thinking. It is a redistribution of thinking.

The redistribution, however, is asymmetric. Verification requires enough task understanding to recognize that checking is needed and enough evidence to judge whether an answer should be accepted. Lee et al. (2025) found that greater confidence in GenAI was associated with less reported critical-thinking effort, while greater task-specific self-confidence was associated with more critical engagement. Participants also described difficulty checking outputs in domains where they lacked expertise. Related human-AI decision research shows that interface transparency alone does not guarantee appropriate reliance. Bansal et al. (2021) found that explanations increased acceptance of AI recommendations regardless of whether those recommendations were correct and did not improve complementary team performance. In an experiment with 199 participants, Bućinca et al. (2021) found that cognitive forcing functions reduced overreliance relative to simpler explainable-AI interfaces, although users rated the more forcing designs less favorably. Klingbeil et al. (2024) similarly showed that an AI label could increase reliance even when advice was disadvantageous. The risk is therefore not merely that AI supplies an answer, but that fluent assistance can arrive before users have engaged sufficiently to evaluate it.

GenAI also reduces the friction that historically limited offloading. Searching several documents, consulting an expert, or constructing a spreadsheet formula requires time and explicit steps. A conversational model can compress those steps into one request. Lower transaction cost makes repeated delegation more likely, especially under time pressure. Yet frequency alone should not be equated with dependence. A professional may use AI repeatedly for low-level drafting while retaining full competence in diagnosis, evaluation, and decision making. Conversely, relatively infrequent use can still be problematic if the user delegates a core learning operation at the exact moment when independent practice is required. The object of analysis should therefore be the cognitive function delegated, not simply the number of prompts sent.

3.3 Immediate performance gains are real, but they are not the same as learning

The strongest positive evidence for GenAI concerns assisted performance. Deng et al. (2025), synthesizing 69 experimental articles, reported that ChatGPT interventions improved academic performance and several affective-motivational outcomes and reduced mental effort. Chen and Cheung

(2025) similarly found positive pooled effects across 57 university studies, including academic achievement and higher-order thinking, although the metacognitive effect was not statistically significant. These findings rule out a blanket claim that AI support necessarily makes users perform worse or think less. In many settings, it helps people produce better work.

Short-term behavioral studies show the same pattern. Jain et al. (2026) compared young adults completing a logical-fallacy task with or without AI assistance; 124 individuals were initially considered and the final analyzed sample was 120 after outlier exclusion. The AI-assisted group produced more correct responses and reported less mental effort. The accompanying qualitative data were more revealing than a simple accuracy comparison: some participants delegated heavily, whereas others used the system strategically and monitored its output. Essel et al. (2024) likewise found higher post-intervention scores on critical, creative, and reflective-thinking scales among undergraduates who used ChatGPT in a research-methods course than among controls using traditional resources. The appropriate conclusion is conditional. GenAI can improve performance and can coexist with cognitively active learning. These studies do not establish that every performance gain becomes independent skill.

Soderstrom and Bjork's (2015) performance-learning distinction explains why the difference matters. If the target outcome is a finished product produced with permanent AI access, assisted performance may be the relevant criterion. If the target is internal mastery, professional resilience when systems fail, examination performance without tools, or transfer to a novel problem, the criterion changes. A system that optimizes the product while bypassing the user's retrieval, generation, or error-correction practice may improve the visible output and still produce little durable change in the person.

3.4 Learning, memory, and retention: evidence for a performance-learning dissociation

Recent experiments provide direct evidence of this dissociation. Bastani et al. (2025) conducted a field experiment with nearly 1,000 high-school mathematics students using two GPT-4-based tutoring conditions. Access to AI improved performance during practice. Yet students given relatively unrestricted assistance later performed worse than controls when they had to solve problems without AI. A more constrained tutor, designed to give teacher-informed hints rather than simply reveal solutions, largely mitigated the learning penalty. The result is difficult to reconcile with either extreme position. AI was clearly useful during supported work, but design choices determined whether that utility translated into unaided learning.

Barcaui (2025) reported a complementary randomized controlled trial with 120 undergraduates. Students used either ChatGPT or traditional non-AI methods while learning content, then completed a surprise test 45 days later. The published study reported lower retention in the ChatGPT-assisted group than in the traditional-study group, 57.5% versus 68.5%, with a medium-to-large standardized difference. The result should not be generalized to every AI-supported learning arrangement, and the delayed-test analysis involved fewer observations than the number originally randomized. Even with those qualifications, the study is valuable because it measures what many AI evaluations omit: what remains after the immediate support period.

The older offloading literature offers plausible mechanisms without licensing causal overreach. If users expect an external system to remain available, they may invest less effort in encoding or rehearsing information (Kelly & Risko, 2022). If AI generates the intermediate steps of a solution, users may also lose opportunities for retrieval practice, error generation, self-explanation, and corrective feedback. The evidence reviewed here does not show that every instance of reduced effort is harmful. Rather, it suggests that effort has to be classified. Eliminating extraneous effort can be beneficial; eliminating the constructive effort that produces a retrievable mental representation can weaken later performance.

3.5 Critical thinking is redistributed, and its quality depends on calibration

Claims that GenAI simply "eliminates critical thinking" are too broad for the available evidence. Lee et al. (2025) found that knowledge workers reported critical-thinking activity in roughly three fifths of their 936 GenAI use examples. The form of critical thinking changed: participants described checking sources, assessing output quality, integrating relevant portions, and adapting generated material to context. This is

genuine cognitive work. At the same time, higher confidence in the AI was associated with lower critical-thinking effort. The study therefore suggests a calibration problem rather than a universal disappearance of reasoning.

Calibration is central because generative outputs can be fluent even when incomplete or wrong. Fisher et al. (2015) demonstrated an analogous metacognitive problem with Internet search: access to external explanations inflated estimates of internal knowledge. The broader automation literature has long distinguished appropriate use from misuse caused by overreliance (Parasuraman & Riley, 1997), and a recent systematic review shows that appropriate trust in AI remains conceptually and operationally difficult to establish across tasks (Mehrotra et al., 2024). Experimental human-AI research further cautions against equating explanation with calibration. Bansal et al. (2021) found that explanations could increase acceptance of both correct and incorrect AI advice, whereas Bućinca et al. (2021) showed that requiring more deliberate engagement could reduce overreliance. The implication is not that users should distrust GenAI by default, but that trust should track evidence and system reliability rather than fluency or interface cues alone.

The behavioral distinction observed by Jain et al. (2026) is therefore useful. AI-assisted participants were not uniformly passive. Strategic users verified, compared, and monitored; others delegated the reasoning process almost entirely. This suggests that critical thinking with GenAI should be measured as a sequence of observable operations, such as problem representation, initial hypothesis generation, source checking, counterexample search, confidence updating, and final justification. A single global question about whether "AI reduces critical thinking" collapses these different processes and makes contradictory findings almost inevitable.

3.6 Creativity and writing: assistance can raise output quality without guaranteeing transfer

Creative work offers one of the clearest demonstrations that assisted performance and independent capability can diverge. Wong and Qiu (2026) randomly assigned 196 university students to solve a creative product-improvement task independently, with free ChatGPT access, or with a regulated "think first, ChatGPT later" approach. The free-AI group benefited on the assisted task. When all participants subsequently completed a more complex creative task without ChatGPT, however, the free-AI group's advantage did not persist. The regulated-AI group, which first generated its own ideas and then used ChatGPT to improve and evaluate them, showed stronger later independent originality.

Process data strengthen the interpretation. In the regulated condition, 88.6% of prompts were collaborative, compared with 29.1% in the free-AI condition. In the free-AI condition, 59.6% of prompts directly requested ideas rather than developing the user's own. Greater use of prompts that improved self-generated ideas was associated with subsequent independent originality and statistically mediated part of the regulated condition's advantage (Wong & Qiu, 2026). The study does not prove that one prompting protocol is universally optimal. It does provide unusually direct evidence that preserving self-generation can matter for transfer.

This pattern helps reconcile positive intervention studies with concerns about dependence. AI feedback, revision support, examples, and alternative formulations can enlarge the learner's representational space. Direct generation can also produce a higher-quality product quickly. The difference is whether the learner must still make consequential cognitive moves. When users generate first, explain why an AI revision is better or worse, compare alternatives, and make the final selection, the system functions more like a scaffold. When they request a finished answer and accept it without reconstruction, the tool substitutes for practice. The product may look similar even though the learning process is not.

3.7 From offloading to dependence

Cognitive offloading and AI dependence should not be treated as synonyms. Offloading is a task strategy. Dependence implies a more persistent reduction in autonomous functioning, such as feeling unable to proceed without the system, repeatedly preferring delegation when independent performance is required, or continuing reliance despite recognized costs. The current dependence literature is developing rapidly,

but much of it remains cross-sectional and self-reported. That limitation should constrain the strength of claims made from it.

Zhang et al. (2024) surveyed 300 university students and modeled relationships among academic self-efficacy, academic stress, performance expectations, and AI dependency. Academic stress and performance expectations were implicated in the pathways to dependency, and participants identified perceived consequences including reduced independent thinking and creativity. Liu et al. (2026), using a separate cross-sectional student sample and structural modeling, likewise found that academic stress was positively associated with GenAI dependency, with performance expectations, academic anxiety, and desire thinking contributing to indirect pathways. Wang et al. (2026) reported associations among AI dependence, academic stress, burnout, anxiety, and lower self-efficacy. These studies identify psychologically coherent correlates, but statistical mediation in cross-sectional data does not establish temporal mediation or show that AI use caused the outcomes.

Qualitative evidence adds phenomenological detail. Farhat (2026), interviewing 45 regular ChatGPT users, documented reports of weaker memory, reduced critical engagement, diminished autonomy, and cognitive reliance, alongside examples of users who deliberately used ChatGPT as a scaffold for reflection and creativity. Such reports are important because dependence is partly experiential, but they cannot quantify objective cognitive change. The strongest conclusion available in 2026 is therefore narrower than popular discourse often implies: problematic reliance exists as a meaningful behavioral and psychological pattern, it is associated with stress and self-regulatory variables, and some users perceive reduced autonomy or cognitive engagement. Evidence for irreversible, generalized decline in human cognitive capacity is not currently established.

4. Discussion

The literature is best understood as a problem of cognitive allocation, not a referendum on whether AI is beneficial or harmful. Across both classic offloading research and GenAI studies, the immediate advantage of external support is consistent: performance can improve because the environment carries part of the computational burden. The cost, when it appears, is usually revealed later, when internal memory, independent transfer, or unaided problem solving is tested. This pattern suggests that the most useful unit of analysis is the cognitive operation displaced by AI and the learning value of performing that operation oneself.

4.1 A substitution-versus-scaffolding account

A substitution-versus-scaffolding account explains much of the apparent inconsistency in the findings. In substitution, AI performs a cognitively productive operation in place of the user. The learner asks for the solution before attempting the problem, requests ideas before generating any, accepts a summary without reconstructing the argument, or uses a recommendation without checking evidence. Immediate performance may improve because the task has become easier. The potential long-term cost is reduced practice in the operation that has been removed.

In scaffolding, AI changes the difficulty or support structure without eliminating the core operation. The user attempts retrieval before asking for a hint, generates an initial explanation before requesting critique, evaluates multiple AI alternatives against explicit criteria, checks cited sources, explains disagreements, or uses AI feedback to revise a self-generated solution. The regulated tutor in Bastani et al. (2025) and the "think first, ChatGPT later" intervention in Wong and Qiu (2026) are different implementations of this principle. Both constrain direct substitution and preserve more human work in the learning loop.

This distinction is related to, but not identical with, established work on automation misuse, overreliance, and appropriate trust. Parasuraman and Riley (1997) framed overreliance as misuse of automation, while contemporary human-AI research asks whether users appropriately accept correct advice and reject incorrect advice (Mehrotra et al., 2024). The present account adds a learning and transfer dimension. A user can appropriately accept a correct AI answer for an immediate task and still lose an opportunity to practice the operation needed for later independent performance. Conversely, evaluating or correcting an

AI output can itself constitute productive cognitive work. The relevant outcome is therefore not only decision correctness with AI, but also what cognitive capability remains after assistance is removed. This account also explains why mental-effort findings should not be interpreted mechanically. Deng et al. (2025) found that ChatGPT reduced mental effort while the same meta-analysis reported positive effects on several learning outcomes. Reduced effort can indicate efficiency when the removed demand is extraneous. It can indicate lost learning opportunity when the removed demand is integral to the target skill. The key question is not "Was the task easier?" but "What became easier, and was that cognitive process itself an instructional target?"

4.2 Four mechanisms that may connect AI offloading to human cognitive outcomes

First, reduced generative processing is a plausible mechanism for weaker retention. Established learning research shows that active retrieval improves delayed retention relative to repeated study (Roediger & Karpicke, 2006), that prompted self-explanation can improve conceptual understanding (Chi et al., 1994), and that generative learning involves actively selecting, organizing, and integrating information with prior knowledge (Fiorella & Mayer, 2016). These literatures do not show that GenAI necessarily impairs learning. They identify cognitive operations that can be displaced when a complete explanation or solution is supplied before the learner attempts the task. Kelly and Risko (2022) show that expected external access can reduce controlled encoding effort, while Barcaui (2025) and Bastani et al. (2025) provide GenAI-specific evidence that easier supported performance can coexist with weaker later unaided performance.

Second, practice displacement is a related but distinct mechanism. Durable competence depends partly on repeated execution of the operations that must later be performed independently. Retrieval research demonstrates that successful exposure during study is not equivalent to practicing retrieval itself (Roediger & Karpicke, 2006). If AI repeatedly performs idea generation, algebraic transformation, source synthesis, or argument construction, users can accumulate finished products without equivalent practice in those operations. Wong and Qiu (2026) provide direct support in the creative domain: participants who first generated their own ideas and then collaborated with AI showed stronger later independent originality than participants with unrestricted AI access.

Third, metacognitive calibration can become distorted. Users must estimate both their own competence and the system's reliability. Offloading decisions are already known to depend on metacognitive confidence (Gilbert et al., 2020). The automation literature shows that overreliance can produce monitoring failures and decision biases (Parasuraman & Riley, 1997), while current human-AI work emphasizes appropriate reliance rather than maximal trust (Mehrotra et al., 2024). Lee et al. (2025) found that higher confidence in GenAI predicted less reported critical-thinking effort, Klingbeil et al. (2024) showed that an AI label could increase reliance, and Buçinca et al. (2021) demonstrated that cognitive forcing can reduce overreliance. Confidence calibration is therefore a cognitive outcome in its own right, not merely a user-experience variable.

Fourth, verification can create a new cognitive bottleneck. GenAI can reduce effort at the production stage while transferring responsibility to evaluation. Whether expertise reliably protects users is not yet established. Domain knowledge may support plausibility checking, source evaluation, and error detection, but older automation research and contemporary human-AI studies caution that competent users can still follow erroneous automated advice. Bansal et al. (2021), for example, found that explanations increased acceptance without selectively improving acceptance of correct advice. Buçinca et al. (2021) showed that deliberately slowing or constraining the interaction could reduce such overreliance. GenAI can therefore move difficult cognition downstream rather than eliminate it; the workload is genuinely reduced only when the remaining evaluative work can be performed accurately.

4.3 Moderators: when the same tool produces different cognitive consequences

Task structure is an obvious moderator. Offloading a date, formatting a citation, or producing a first-pass transcription carries different learning stakes from offloading diagnostic reasoning, statistical interpretation, conceptual explanation, or the derivation of a mathematical procedure. Tasks whose

educational purpose is to practice a cognitive operation are especially vulnerable to substitution. In contrast, professional tasks in which the goal is an accurate external product may rationally prioritize augmented performance, provided that verification and accountability remain intact.

Prior knowledge is likely to moderate the balance, but direct GenAI evidence is still limited. Domain knowledge can plausibly support prompt formulation, source evaluation, and error detection, while Lee et al. (2025) found that greater task-specific self-confidence was associated with more reported critical-thinking engagement. At the same time, human-AI studies show that explanation alone does not guarantee appropriate reliance (Bansal et al., 2021) and that more effortful decision procedures can sometimes reduce overreliance (Bućinca et al., 2021). The commonly assumed novice-expert gradient should therefore be treated as an empirical question rather than a settled rule. Future studies should test whether expertise changes when users delegate, verify, or reject AI-generated content and whether those differences predict later independent performance.

Trust and self-confidence are supported moderators, whereas time pressure remains a plausible but insufficiently tested factor in the GenAI literature reviewed here. Lee et al. (2025) reported opposing associations of AI confidence and task-specific self-confidence with critical-thinking engagement. Zhang et al. (2024) and Liu et al. (2026) linked academic stress and performance expectations with dependency constructs. These studies support an association between stress-related variables and AI dependence, but their cross-sectional designs do not establish that time pressure causes a transition from occasional offloading to habitual reliance. That pathway is theoretically plausible and should be tested longitudinally rather than stated as an established mechanism.

Finally, interface and instructional design matter. Bastani et al. (2025) showed that guardrails can substantially change learning outcomes even when the underlying generative model is similar. Wong and Qiu (2026) showed that sequencing the interaction so that users think first can change both prompting behavior and later independent performance. These findings shift part of the responsibility from individual self-control to system architecture. Tools can be designed to ask for a learner's attempt, offer graduated hints, request justification, expose uncertainty, require source checks, and periodically test unaided performance.

4.4 Implications for education

Educational policy should distinguish between AI for production and AI for learning. If the objective is independent reasoning, unrestricted answer generation during practice can be counterproductive even when it improves supported work, as the PNAS mathematics experiment demonstrates (Bastani et al., 2025). Established learning research also shows that retrieval practice and self-explanation can contribute to durable retention and conceptual understanding (Chi et al., 1994; Roediger & Karpicke, 2006). A defensible educational design, therefore, preserves phases of unaided retrieval or generation when those operations are themselves learning objectives, uses AI after an initial attempt or as graduated support, and periodically assesses students without AI so that independent mastery can be distinguished from tool-assisted performance.

This does not require banning GenAI. A stronger approach is to specify which operations may be offloaded at each stage. Early learning might restrict direct solution generation while allowing hints, examples, Socratic questioning, or feedback on an attempted solution. Later, when competence is established, broader AI use can support speed, comparison, and professional workflow. Assessment can likewise separate "with-AI performance" from "independent mastery" rather than pretending that one measure captures both.

Students also need metacognitive AI literacy. This includes distinguishing confidence from evidence, stating an initial judgment before consulting AI when appropriate, checking high-stakes claims against primary or authoritative sources, and explaining why an AI suggestion was accepted, rejected, or revised. Research on appropriate trust shows why simple exposure to explanations is insufficient: explanations can increase acceptance without improving selectivity (Bansal et al., 2021), whereas cognitive forcing can reduce overreliance by requiring more deliberate engagement (Bućinca et al., 2021). The educational

objective is not ceremonial fact-checking. It is calibrated reliance in which epistemic responsibility remains with the learner.

4.5 Implications for knowledge work and professional practice

In professional settings, the appropriate benchmark differs from education because permanent tool access may be part of the job. Even there, a purely output-based evaluation is incomplete. Organizations need enough internal competence to recognize model error, handle outages, manage unusual cases, and supervise automated workflows. Overreliance becomes an operational risk when the person responsible for a decision cannot reconstruct the reasoning or detect a failure.

Workflows can therefore be stratified by cognitive criticality. Low-stakes, reversible tasks such as formatting, routine rewriting, and first-pass summarization can tolerate greater substitution. High-stakes or expertise-defining tasks should retain stronger human checkpoints, such as an independent initial judgment, explicit uncertainty estimates, source verification, or a rationale that can be defended without simply reproducing model output. Lee et al. (2025) suggest that GenAI already shifts some knowledge work toward verification, integration, and stewardship. Evidence from human-AI decision-making further shows that explanations do not automatically produce appropriate reliance (Bansal et al., 2021). Organizations should therefore treat verification and calibrated reliance as trainable competencies rather than assuming they emerge automatically from tool use.

5. Limitations of the Evidence and Research Agenda

The most important limitation in the current literature is timescale. GenAI has been widely available for only a few years, and many studies examine a single session, a short course, or self-reported habits. Strong claims about chronic cognitive decline, permanent memory change, or population-level effects therefore exceed the available evidence. Longitudinal studies are needed that measure baseline ability, detailed AI-use behavior, and independent cognitive outcomes over months or years.

A second limitation is construct validity. "Critical thinking" is often measured by self-report, course-specific scales, or assisted task quality. These are not interchangeable with objective reasoning performance. Future studies should use blinded scoring, validated reasoning tasks, delayed tests, transfer problems, and process measures that capture what users actually do before, during, and after consulting AI. The same applies to dependence. Frequency of use, preference for convenience, overtrust, compulsive use, and loss of independent competence are related but distinct constructs and should not be collapsed into one score.

Third, future experiments should manipulate the cognitive operation delegated rather than only comparing "AI" with "no AI." Useful conditions include answer generation versus hints, AI-first versus human-first sequencing, source-visible versus source-hidden outputs, mandatory verification versus optional verification, and continuous assistance versus strategically faded assistance. Such designs can identify which elements of AI interaction cause performance gains and which support or undermine learning.

Fourth, individual differences require closer study. Working-memory capacity, prior knowledge, self-efficacy, age, motivation, and metacognitive calibration may all influence when offloading is adaptive. Burnett and Richmond's (2026) meta-analysis suggests that external supports can reduce performance disparities in memory tasks, an important potential benefit. The challenge is to obtain that compensatory benefit without preventing the development of internal skill where internal skill remains necessary.

Finally, research should report two outcomes whenever possible: assisted performance and subsequent unaided performance. The difference between them is theoretically informative. A large assisted gain followed by no independent gain suggests augmentation without learning. A smaller assisted gain followed by stronger independent performance may indicate productive scaffolding. Without both measures, studies risk mistaking a better artifact for a better-trained mind.

6. Conclusion

Generative AI does not create a new human tendency to externalize cognition. It expands the range of cognition that can be externalized and sharply reduces the cost of doing so. The empirical record shows

why this distinction matters. Cognitive offloading can improve immediate performance and release resources for other work. GenAI can likewise improve academic output, supported reasoning accuracy, creativity, and efficiency. These benefits are supported across multiple study designs, although their size and durability vary substantially by task and intervention.

The evidence also identifies an important boundary condition. When AI substitutes for operations that a person needs to learn, assisted success can overstate internal learning. Recent experiments converge on a narrower and defensible lesson: unrestricted direct-answer assistance can improve supported performance without guaranteeing later independent performance, whereas conditions that preserve self-generation or constrain direct substitution can better protect transfer in some contexts (Barcaui, 2025; Bastani et al., 2025; Wong & Qiu, 2026). The specific effects of verification, explanation, active revision, and support fading remain less directly isolated experimentally and should not be treated as equally established mechanisms.

The most defensible answer to the title question is therefore conditional. People can think with AI and, in some circumstances, think less because of AI. The difference is not captured by usage frequency alone. It depends on what cognitive work is delegated, whether the user can evaluate the output, and whether the interaction preserves the mental operations needed for future independent performance. The research priority is no longer to ask whether GenAI is simply good or bad for cognition. It is to identify forms of human-AI collaboration that improve present performance without eroding the capabilities required for future independent judgment.

Acknowledgements

The authors gratefully acknowledge the support of the Research Division, KIU, and the Faculty of Behavioural Sciences, KIU, for their institutional support in facilitating this scholarly work.

Conflict of Interest

The authors declare no competing interests.

Funding

No external funding was received for this review.

Ethics Statement

Not applicable. This narrative review did not involve the collection of new data from human participants or animals

.

Data Availability

No new dataset was generated or analyzed for this narrative review.

References

- Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA: A scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*, 4, 5. <https://doi.org/10.1186/s41073-019-0064-8>
- Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. S. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Article 81, pp. 1–16). Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445717>
- Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open*, 12, 102287. <https://doi.org/10.1016/j.ssaho.2025.102287>
- 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*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–21. <https://doi.org/10.1145/3449287>
 - Burnett, L. K., & Richmond, L. L. (2026). Meta-analytic investigations of the effect of cognitive offloading on memory-based task performance and interindividual variability. *Memory & Cognition*, 54(1), 144–168. <https://doi.org/10.3758/s13421-025-01743-8>
 - Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis. *Educational Research Review*, 49, 100737. <https://doi.org/10.1016/j.edurev.2025.100737>
 - Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3
 - Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
 - Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, 6, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
 - Farhat, Z. (2026). When AI thinks for us: Unveiling the cognitive and social toll of ChatGPT overreliance. *Applied Cognitive Psychology*, 40(3), e70205. <https://doi.org/10.1002/acp.70205>
 - Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717–741. <https://doi.org/10.1007/s10648-015-9348-9>
 - Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, 144(3), 674–687. <https://doi.org/10.1037/xge0000070>
 - Gilbert, S. J. (2015). Strategic offloading of delayed intentions into the external environment. *Quarterly Journal of Experimental Psychology*, 68(5), 971–992. <https://doi.org/10.1080/17470218.2014.972963>
 - Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2020). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General*, 149(3), 501–517. <https://doi.org/10.1037/xge0000652>
 - Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, 74(9), 1477–1496. <https://doi.org/10.1177/17470218211008060>
 - Jain, P., Shajit, D., Maini, A., Shahin, Y., & Panda, A. (2026). Cognitive offloading, critical thinking and attitudes towards artificial intelligence in the era of ChatGPT: A comparative study of artificial intelligence-assisted and manual task performance in young adults. *Cognitive Processing*. Advance online publication. <https://doi.org/10.1007/s10339-026-01375-z>
 - Kelly, M. O., & Risko, E. F. (2022). Study effort and the memory cost of external store availability. *Cognition*, 228, 105228. <https://doi.org/10.1016/j.cognition.2022.105228>
 - Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI: An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, 160, 108352. <https://doi.org/10.1016/j.chb.2024.108352>
 - Lee, H.-P. (H.), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Article 1121, pp. 1–22). Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713778>

- Liu, X., Liu, Y., Dai, Y., & Fu, J. (2026). Academic stress and university students' dependency on generative artificial intelligence: A multiple mediation model using PLS-SEM. *BMC Psychology*, 14, 216. <https://doi.org/10.1186/s40359-026-03986-9>
- Lu, X., Kelly, M. O., & Risko, E. F. (2020). Offloading information to an external store increases false recall. *Cognition*, 205, 104428. <https://doi.org/10.1016/j.cognition.2020.104428>
- Mehrotra, S., Degachi, C., Vereschak, O., Jonker, C. M., & Tielman, M. L. (2024). A systematic review on fostering appropriate trust in human-AI interaction: Trends, opportunities and challenges. *ACM Journal on Responsible Computing*, 1(4), Article 26, 1–45. <https://doi.org/10.1145/3696449>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199. <https://doi.org/10.1177/1745691615569000>
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776–778. <https://doi.org/10.1126/science.1207745>
- Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological Science*, 26(2), 182–188. <https://doi.org/10.1177/0956797614559285>
- Wang, W., Wu, Y., Fang, J., Yang, C., & Wen, L. (2026). When cognitive offloading becomes dependence: How AI dependence mediates the pathway from academic stress to burnout and anxiety. *BMC Psychology*, 14, 1125. <https://doi.org/10.1186/s40359-026-04911-w>
- Wong, S. S. H., & Qiu, S. X. (2026). Think first, ChatGPT later: Guiding human-AI collaboration for learning gains in independent human creativity. *Educational Psychology Review*, 38, 45. <https://doi.org/10.1007/s10648-026-10118-7>
- Zhang, S., Zhao, X., Zhou, T., & Kim, J. H. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. *International Journal of Educational Technology in Higher Education*, 21, 34. <https://doi.org/10.1186/s41239-024-00467-0>