International Journal of Multidisciplinary and Current Educational Research (IJMCER)
ISSN: 2581-7027 ||Volume|| 7 ||Issue|| 5 ||Pages 01-10 ||2025||

Working Memory in the Age of Artificial Intelligence: Cognitive Paradoxes and Educational Implications

Sacide Güzin Mazman Akar
Department of Instructional Technologies, Usak University. 64200 Usak, Türkiye.
ORCID: https://orcid.org/0000-0003-2188-221X

ABSTRACT : This paper moves beyond the "use AI or not" debate and treats AI–learner interaction as co-regulation of working memory. Drawing on Cognitive Load Theory, retrieval practice, and metacognition, it outlines when AI helps learning and when it hurts. AI helps when it reduces needless external load, breaks complex tasks into steps, and channels effort toward building schemas. It hurts when verbose or misleading outputs overload working memory or replace retrieval. Two levers matter most: when help is given and how detailed it is. A practical routine is: learners first try on their own, then receive a short, local hint, and later face a delayed check. Design tips include using step-by-step, local explanations; "skeleton-then-expand" for writing; fading support over time; and matching format to the discipline. The paper proposes a measurement agenda combining three-part cognitive load scales, process traces (hints, steps, time), and delayed recall/transfer tests. Limits and boundary conditions are noted (e.g., LLMs are not human working memory). Overall, AI should act as a scaffold, not a crutch: keep help brief and after the learner's attempt, protect retrieval opportunities, and limit unnecessary verbosity.

KEYWORDS: working memory; cognitive load theory; generative AI; scaffolding; retrieval practice; instructional design; cognitive offloading; educational technology

I. INTRODUCTION

Large language models (LLMs) and related AI tools increasingly function as dynamic external memory and strategy providers, yet empirical results oscillate between clear benefits and worrisome dependency. To reconcile these findings, Cognitive Load Theory is integrated with accounts of cognitive offloading and metacognitive control to propose an AI–Learner Working-Memory Co-Regulation Framework. The framework specifies how design routines and safeguards align assistance with the learner's working-memory (WM) load profile, predicting when AI yields efficient encoding and transfer versus extraneous-load inflation and weak generalization. The contribution is threefold: (i) a unified theory linking assistance timing and output granularity to WM components and outcomes; (ii) a design agenda that operationalizes supportive versus strain pathways for STEM problem solving and academic writing; and (iii) a measurement plan pairing tri-component load instruments with trace-based analytics and delayed learning tests. Building on this synthesis, a measurement-and-design agenda (load instruments, traces, delayed outcomes) is outlined to advance cumulative, mechanism-oriented research without formal hypotheses. Neurocomputational work outlines a three-level architecture—sensorimotor (local, nonconscious), a global but nonconscious cognitive level, and a GNW-based conscious level—that specifies minimal requirements for developing cognitive abilities (Volzhenin et al., 2022). The model underscores that adapting neural connections at local and global levels is required to solve progressively complex tasks.

Building on cognitive load theory and load-reduction instruction, prior mechanisms that reduce extraneous load (e.g., pretraining, signaling, example-first sequencing) and subsequently fade guidance as expertise develops are acknowledged. Classical worked-example research has primarily optimized what and how much information is provided before independent problem solving (Barbieri et al., 2023). In contrast, the proposed framework treats the timing of assistance and the granularity of outputs as the primary design levers through which working memory is co-regulated. This shift moves beyond a general "more-versus-less guidance" stance: assistance delivered too early and at high granularity can suppress retrieval opportunities and impair delayed transfer, even while lowering immediate effort; whereas later, minimal, local assistance can preserve retrieval while constraining extraneous load (Endres et al., 2024). Guidance is thus reconceptualized not merely as a quantity to be faded, but as a temporal–granular control problem with predictable, testable dissociations in effort, accuracy, and long-term learning.Recent meta-analytic reviews of AI-assisted instruction typically report small-to-moderate average benefits accompanied by substantial heterogeneity (Li et al., 2025; Alanazi et al., 2025). A major source of this variance is proposed to be the under-specification of assistance timing and output

Granularity—an extension of the long-standing "assistance dilemma"—with recent syntheses also showing that instructional context (e.g., guided vs. unguided use) systematically moderates effects (Qu et al., 2025; Koedinger & Aleven, 2007). By explicitly modeling these two levers, it is predicted that AI support will help or hinder in systematic ways: early, high-granularity outputs are likely to inflate immediate performance yet depress later recall/transfer by crowding out retrieval; delayed, minimal, and locally targeted prompts should yield more durable learning (Roediger & Butler, 2011; Pastötter & Frings, 2019). Accordingly, the question is reframed from "Does AI help?" to "Under which timing-by-granularity configurations does AI help, for whom, and on which outcomes (immediate vs. delayed; near vs. far transfer)?" In this view, mixed effects are explained and design-testable moderators for future trials and meta-analytic coding are furnished (Wang & Fan, 2025; Alanazi et al., 2025).Artificial intelligence (AI) technologies—ranging from conversational agents (e.g., ChatGPT) to adaptive tutoring systems and memory-augmented architectures—are increasingly embedded in formal and informal learning. Beyond broad societal uptake, education has become a primary arena for AI's promises of personalization, formative feedback, and workload reduction (Roy et al., 2024; Miao & Holmes, 2023; OECD, 2023). Policy and synthesis reports highlight AI's potential to strengthen feedback loops, intensify formative assessment, and tailor instruction when thoughtfully implemented (Cardona et al., 2023). Empirical syntheses similarly report positive mean effects of AI-supported tools on achievement and higher-order outcomes, including recent meta-analyses on ChatGPT-assisted learning and intelligent tutoring in K–12 (Wang & Fan, 2025; Dong et al., 2025; Létourneau et al., 2025).

At the same time, the integration of AI foregrounds working-memory constraints and cognitive load. CLT-aligned proposals and reviews argue that generative AI can optimize load—e.g., reducing extraneous processing or adaptively structuring materials—yet may also inject fragmented, task-irrelevant information that inflates load when poorly designed (Gkintoni et al., 2025; Martin et al., 2025; Twabu, 2025). This ambivalence mirrors findings on cognitive offloading—delegating mental operations to external tools—which can free resources for reasoning but also divert effort from internal encoding and retrieval (Risko & Gilbert, 2016). In education, such dynamics now extend from search/note-taking to LLM-based assistants that draft, summarize, and explain on demand.Evidence on overreliance costs is accumulating. In higher-education samples, overuse of ChatGPT has been associated with procrastination, memory loss, and dampened performance, consistent with concerns that reliance on AI may erode durable knowledge if it replaces rather than scaffolds retrieval-based learning (Abbas et al., 2024). Experimental work from the MIT Media Lab reports reduced neural engagement and poorer memory/ownership of produced text during LLM-assisted writing compared to search-assisted or unaided writing—implicating motivational and metacognitive mechanisms alongside WM (Kosmyna et al., 2025). These patterns echo the "Google effect": easy access to information encourages remembering where to find facts rather than the facts themselves—i.e., a shift toward transactive memory that can trade off with internal storage unless instruction deliberately re-engages retrieval (Sparrow et al., 2011).

Crucially, these dynamics are not uniform across ages and disciplines. Grade-level comparisons suggest developmental differences in offloading strategies: younger learners may underuse optimal externalization, whereas adolescents may over-rely on reminders—pointing to age-appropriate metacognitive instruction on when and how to offload (Sun et al., 2025). In higher-education and discipline-specific contexts, early studies of course-embedded AI chatbots indicate performance benefits while also documenting shifts in intrinsic/extraneous load that depend on task design and the granularity of prompts and explanations (Lademann et al., 2025). Taken together, these findings suggest that AI's educational value hinges on co-regulating learners' WM demands: leveraging AI to scaffold schema construction and reduce extraneous load while preserving desirable difficulties (e.g., retrieval practice) that consolidate long-term memory. Working-memory limits and typical AI functions are first situated within an educational frame (Section 2; see Fig. 1). Evidence on supportive mechanisms (Section 3) and strain mechanisms (Section 4) is then synthesized, and the resulting insights are distilled into design principles for educators and developers (Section 5). Finally, a future research agenda is outlined that details measurement and task-design guidance (Section 6), without advancing formal hypotheses.

II. WORKING MEMORY AND ARTIFICIAL INTELLIGENCE IN EDUCATION: A CONCEPTUAL FRAME

Working memory (WM) is the capacity-limited system that temporarily holds and manipulates information during learning and problem solving. In instructional contexts, WM burden comprises intrinsic demands from task complexity, extraneous demands from presentation/interface factors (e.g., fragmented displays), and germane effort invested in schema construction. Designs that minimize extraneous processing and stage intrinsic complexity yield more durable understanding and transfer, whereas multitasking and rapid context switching inflate load and impair recall and comprehension (Twabu, 2025; Grinschgl et al., 2021). Durable learning also requires episodic encoding and retrieval practice; when intermediate steps are routinely externalized, learners encode fewer episodic details and engage in fewer retrieval events, weakening later recall/transfer (Sparrow et al., 2011; Oakley et al., 2025).Contemporary AI tools—including LLM-based assistants—can alleviate WM demands when they personalize content, scaffold steps, and provide concise exemplars or explanations; used well, these functions compress extraneous processing and free WM for sense-making (Cardona et al., 2023). Used poorly, they may short-circuit productive struggle and displace retrieval, producing mixed or negative learning effects (Risko & Gilbert, 2016; Abbas et al., 2024; Wang & Fan, 2025). Importantly, while LLMs maintain local context windows and surface tool-assisted "memory," they do not possess human WM or episodic systems; recent analyses highlight architectural/scaling limits not isomorphic to human WM constraints. Accordingly, educational use should treat AI as a context-sensitive external aid, not a cognitive surrogate (Gong & Zhang, 2024; Huang et al., 2025; Cardona et al., 2023).

A schematic summary of these supportive versus strain pathways appears in Figure 1.

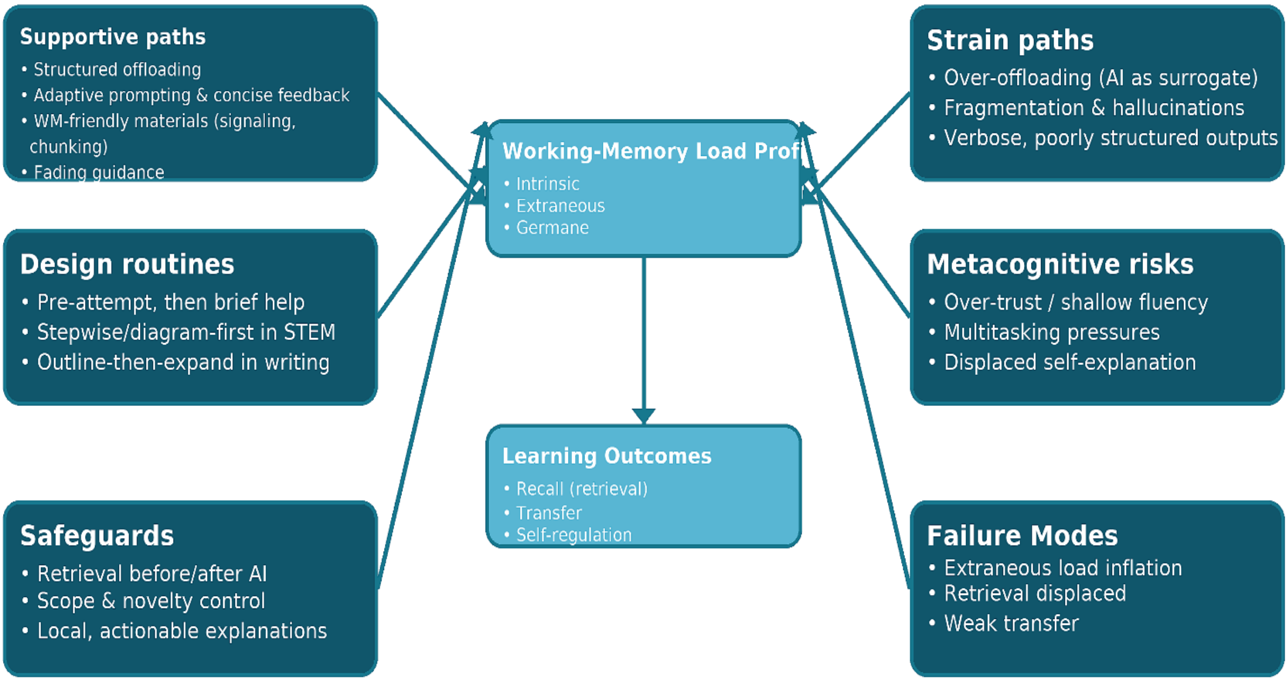

Figure 1. *Supportive versus strain mechanisms by which AI influences working memory (WM) in education Note. Left-hand boxes show supportive paths—structured offloading, adaptive prompting with concise feedback, WM-friendly materials (signaling, chunking), and fading guidance. Right-hand boxes summarize strain paths and metacognitive risks—over-offloading (AI as surrogate), fragmentation/hallucinations, verbose/poorly structured outputs. Arrows indicate how these inputs shape the WM load profile (intrinsic–extraneous–germane) and, in turn, downstream learning outcomes (recall, transfer, self-regulation). Abbreviations: AI = artificial intelligence; WM = working memory.*

Relative to load-reduction instruction and the worked-example tradition, the contribution is not merely the proposal of "more example-first" or "better prompts"; rather, AI–learner interaction is formalized as working-memory co-regulation along two orthogonal axes—assistance timing and output granularity. Through this formalization, concrete, falsifiable contrasts (early vs. late; global vs. minimal/local) are yielded and can be mapped to distinct causal pathways (support vs. strain) and boundary conditions (expertise level, task novelty, abstraction). Positioned in this way, classical guidance-fading insights are integrated with contemporary AI affordances, and meta-analytic heterogeneity is explained by specifying when identical content, delivered at different times and granularities, is likely to produce opposite patterns in immediate performance, cognitive-load profiles, and delayed transfer.

III. SUPPORTIVE MECHANISMS: HOW AI CAN *REDUCE* WORKING-MEMORY BURDEN

AI can be deliberately orchestrated to lower extraneous load, stage intrinsic complexity, and increase germane processing.
(a) Structured cognitive offloading. Breaking complex tasks into prompted sub-steps, providing just-in-time hints, and formatting outputs to reduce split-attention eases WM bottlenecks and supports novices when demands are high (Risko & Gilbert, 2016; Cardona et al., 2023). In writing and problem solving, AI can supply formats, checklists, and exemplars that cut display/search costs while keeping learners engaged in core reasoning (Kosmyna et al., 2025; Lademann et al., 2025).(b) Adaptive sequencing and concise feedback. Diagnosing misunderstandings, adjusting step size, and delivering brief, specific, action-oriented feedback aligns assistance with the learner's momentary WM state (Cardona et al., 2023; Twabu, 2025).(c) WM-friendly materials and faded guidance. Summaries, signaling (headings, bullets), and chunked exemplars curb extraneous search, while worked examples with fading foster schema acquisition without flooding WM; teachers can use AI to generate isomorphic examples with controlled complexity and fade assistance across trials (Twabu, 2025).

IV. STRAIN MECHANISMS: HOW AI CAN *INCREASE* WORKING-MEMORY BURDEN

Persistent reliance on AI to recall facts or complete intermediate steps can displace retrieval, weakening later recall and transfer (Sparrow et al., 2011; Oakley et al., 2025). Mixed or negative learning effects arise when generative tools substitute for—rather than scaffold—core processing (Abbas et al., 2024; Wang & Fan, 2025). Long, verbose, or poorly structured outputs inflate extraneous load; hallucinated or partially relevant content imposes verification/reconciliation costs that further tax attention and WM (Twabu, 2025; Cardona et al., 2023). A metacognitive pathway compounds these burdens: when answers are instantly available, learners may overestimate understanding, under-invest in monitoring/encoding, and drift toward transactive memory strategies (Risko & Gilbert, 2016; Shanmugasundaram & Tamilarasu, 2023). The result can be short-term fluency but long-term fragility, especially when retrieval practice is bypassed or outputs are fragmented/unvetted.

V. DESIGN PRINCIPLES AND TEACHING ROUTINES (AI × EDUCATION × WORKING MEMORY)

Effective use of AI should co-regulate rather than outsource WM.

(1) Load-aware prompting. Constrain scope to one idea or worked step; request headings/numbered steps/brief takeaways to reduce search costs; control novelty by introducing one new element at a time when content is unfamiliar (Twabu, 2025).

(2) Retrieval-preserving workflows. Elicit an initial unaided attempt before any AI exemplar; follow help with short, delayed quizzes/generation to restabilize memory; fade guidance across practice (Oakley et al., 2025).

(3) Level- and discipline-matching. For novices, favor worked examples, tight formatting, and shorter outputs; for advanced learners, reduce scaffolds and invite more open-ended generation. In STEM, prefer stepwise derivations and diagram-first prompts; in writing, use outline-then-expand routines (Cardona et al., 2023).

(4) Explanations that serve WM. Keep explanations brief, local, and actionable (why this step? what to check next?), reserving deeper model-level accounts for cases where they improve decision quality or self-regulation—avoiding verbosity that inflates extraneous load (Cardona et al., 2023).Building on the AI–Learner Working-Memory Co-Regulation Framework and the pathways summarized in Fig. 1, future work should prioritize mechanism-oriented studies that are comparable across settings without requiring formal hypotheses. The shared aim is to make design choices and measurement practices explicit so that findings on AI-supported learning can cumulate coherently around how assistance shapes the WM load profile (intrinsic–extraneous–germane) and, in turn, recall, transfer, and self-regulation (Risko & Gilbert, 2016; Sparrow et al., 2011; Twabu, 2025).

A first priority is to systematically vary the design levers your review identifies as most consequential: the timing of assistance (pre-attempt vs. brief post-attempt help), the granularity of outputs (stepwise/diagram-first and outline-then-expand vs. final/verbose), retrieval safeguards (short unaided attempts before assistance and delayed retrieval after), explanation safeguards (local, actionable rationales tied to specific steps), the structure of offloading (signaling, chunking, checklists), and scope/novelty control (one idea or step per prompt, staged introduction of new elements). These levers should be exercised in discipline-sensitive tasks—stepwise derivations and diagram-first prompts in STEM; outline-then-expand in writing—and across learner levels to surface boundary conditions such as expertise reversal and multitasking pressure (Cardona et al., 2023; Wang & Fan, 2025; Lademann et al., 2025; Abbas et al., 2024).To support cumulation, we recommend a minimal common battery that combines subjective and trace/product measures with delayed outcomes. Immediately after tasks, administer short tri-component cognitive load scales for intrinsic, extraneous, and germane load. Pair these with process traces (prompt/output token counts, step counts, revisions, time-on-step, backtracks) and product analytics (e.g., claim–evidence links in writing; step-validation in STEM). Use immediate and delayed recall (24–72 h) and differentiate near vs. far transfer; include simple manipulation checks (perceived stepwise-ness, verbosity, novelty/scope, trust). This battery makes it possible to map each design decision to predicted load shifts and downstream learning, aligning with your framework's support/strain pathways (Cardona et al., 2023; Sparrow et al., 2011; Twabu, 2025).

Study designs can remain researchable without formal propositions while still being confirmatory in spirit. In STEM, a lab/classroom hybrid can randomize learners to pre- vs. post-attempt help crossed with stepwise vs. final outputs, use isomorphic items for near transfer, and include a delayed recall check; in writing, course-embedded studies can contrast outline-then-expand with final-essay assistance, optionally adding brief local explanations and a short unaided outline before any AI exposure. In both cases, report representative prompt–output pairs, model/version details, and any guardrails used to preserve retrieval or control verbosity (Kosmyna et al., 2025; Lademann et al., 2025).Because effects will vary by learner and context, analyses should model mechanisms and moderators. Report scale reliabilities and manipulation checks; use mediation to test whether extraneous↓ and germane↑ plausibly carry design effects to outcomes; probe moderation by prior knowledge, task complexity, grade level, and time pressure (expertise-reversal, multitasking). Mixed-effects models can account for repeated tasks and classroom clustering. When feasible, preregister core contrasts and share anonymized traces, rubrics, and code to improve reproducibility and synthesis (Risko & Gilbert, 2016).Finally, the design guidance emerging from your synthesis can be distilled for practice and then evaluated empirically: gate help by timing (solicit a brief unaided attempt, then provide concise, targeted assistance and re-engage memory with a short delayed check); keep outputs short and structured (stepwise, diagram-first, outline-then-expand, one new element at a time); scaffold, then fade support as competence grows; and explain locally with actionable rationales rather than long narratives that inflate extraneous load. Framed this way, future work advances a programmatic agenda in which AI serves as a scaffold, not a surrogate, co-regulating WM demands to protect encoding and transfer while respecting developmental and disciplinary differences (Abbas et al., 2024; Wang & Fan, 2025; Sparrow et al., 2011; Twabu, 2025).

VI. DISCUSSION

This paper reframes human–AI interaction as working-memory (WM) co-regulation rather than an "AI vs. no-AI" dichotomy. Grounded in Cognitive Load Theory (CLT), we synthesize a dual pattern: when AI is orchestrated as load-sensitive scaffolding, it reduces extraneous processing, stages intrinsic complexity, and channels effort toward germane processing; when it becomes a surrogate for core cognition, it risks over-offloading and retrieval displacement, yielding short-term fluency but long-term fragility (Chandler & Sweller, 1992; Sweller, 1994; Risko & Gilbert, 2016; Twabu, 2025; Abbas et al., 2024; Wang & Fan, 2025). Our framework treats AI as a designable scaffold tuned to the learner's load profile, not a blanket substitute for core cognitive work.

Mechanism 1: Cost-saving without displacement. When AI surfaces definitions, exemplars, and local cues at the moment of need, it trims search/friction and limits split attention, thereby lowering extraneous load—provided that core generative moves remain with the learner (Risko & Gilbert, 2016; Lademann et al., 2025). Benefits peak when help is short, local, and step-bound, aligning with evidence that worked examples and example–problem sequences aid novices if tied to the focal step and then fade to promote independent problem solving (Barbieri et al., 2023; Renkl & Atkinson, 2016; van Gog et al., 2011).

Mechanism 2: Protecting retrieval opportunities. Durable learning depends on retrieval. Giving answers too early blocks the retrieval that consolidates memory and supports transfer. Use a simple routine: pre-attempt → minimal, local hint → delayed check. Learners try first, get targeted help, then complete a later test. At a global level, dopamine seems necessary for credit assignment across the delay between perception and reward, which supports using delayed checks (Volzhenin et al., 2022). If AI replaces rather than brackets these retrieval moments, later recall and transfer decline (Oakley et al., 2025). Across lab and classroom studies, retrieval (including delayed or repeated testing) reliably outperforms restudy for long-term retention and can even enhance subsequent learning—the forward-testing effect (Karpicke & Roediger, 2008; Roediger et al., 2011; Pastötter & Bäuml, 2014).

Mechanism 3: Granularity and locality of AI output. Long, global explanations or highly creative, off-scope text tend to inflate extraneous load—especially for low-prior-knowledge learners—whereas stepwise, local explanations support schema construction (germane load) without overwhelming WM. CLT's guidance-fading and expertise-reversal effects imply that the same prompt strategy can help novices yet hinder more knowledgeable learners unless assistance is progressively withdrawn and localized to genuine impasses (Sweller et al., 2011; Kalyuga et al., 2003). Long, verbose, or poorly structured responses also impose verification burdens when content is hallucinated or only partially relevant (Twabu, 2025; Cardona et al., 2023).
Contextual moderators: K–12 vs. higher education. Moving from mechanisms to context, K–12 settings—where metacognitive control and selective offloading are still developing—benefit most from worked examples, stepwise prompts, and brief, highly structured outputs, with dense scaffolding that fades across isomorphic items (Rosenshine, 2012; Koedinger & Corbett, 2006; Renkl et al., 2002; CESE, 2017; Barbieri et al., 2023). In higher education, where prior knowledge is greater and expertise-reversal can make heavy guidance counterproductive, AI should act as a local, minimalist aid: require a pre-AI attempt, request brief, local explanations, and fade quickly to promote self-explanation and transfer (Kalyuga et al., 2003; Wittwer & Renkl, 2008; Zawacki-Richter et al., 2019). Finally, ubiquitous multitasking and continuous partial attention act as chronic environmental pressures that erode WM and creative synthesis, complicating attempts to attribute gains or deficits solely to AI (Fırat, 2013).

Offloading vs. ownership. Decades of work on cognitive offloading show that when external stores are reliable, people preferentially remember where to find information rather than the information itself (the "Google effect"). Offloading is rational only if retrieval opportunities are preserved and core generative steps remain with the learner; otherwise, near-term fluency can mask fragile, non-transferable knowledge (Sparrow et al., 2011).

Automation bias as a boundary risk. Fluent AI suggestions can induce automation bias, leading users to overweight AI outputs and under-weight contradictory evidence. In education and clinical training, analyses argue that this bias short-circuits critical appraisal unless interfaces and routines explicitly re-instate verification and reflection—reinforcing our emphasis on structured pre-attempts and delayed checks rather than "answer-first" assistance (Nguyen, 2024; Romeo & Conti, 2025).

Emerging evidence from AI-enhanced instruction. Recent experiments and reviews indicate that well-designed AI tutors and course-embedded chatbots can reduce intrinsic/extraneous load, raise self-efficacy and interest, and in some cases match or surpass strong baselines—while long-term performance and transfer remain uneven across tasks and designs. This pattern is consistent with our claim that assistance timing and scope determine whether AI functions as scaffold or crutch (Lademann et al., 2025; Kestin et al., 2025).

Measurement implications. Rigorous tests of AI–WM co-regulation should combine multidimensional cognitive-load scales (intrinsic, extraneous, germane), Paas mental effort ratings, and trace-based analytics (e.g., hint usage, dwell time, backtracks) with delayed recall and transfer assessments. Validations of Leppink's CLS in online/virtual environments support pairing subjective load with behavioral traces, which learning-analytics studies show can capture engagement tactics at scale (Andersen & Makransky, 2021; Choi & Lee, 2021; Ouwehand et al., 2021).

Design agenda. The framework yields four testable contrasts for classroom experiments and product evaluations: (a) assistance timing (pre-attempt vs. post-attempt minimal hint), (b) granularity (stepwise/local vs. global/long-form output), (c) fading (fixed help vs. progressive silencing of cues), and (d) assessment placement (immediate verification vs. 24–72-hour delayed retrieval). Together with expertise-sensitive routing, these contrasts specify when AI should act as an adaptive scaffold that trims avoidable load while protecting the learner's generative and retrieval work (Kalyuga et al., 2003).
Overall, the discussion integrates classic CLT effects (split attention, worked examples, fading, expertise reversal) with findings on retrieval practice, cognitive offloading, and automation bias to explain why assistance timing and granularity are the pivot points of AI's educational value. Designing routines around short, local help after a learner attempt and delayed retrieval offers a principled path to convert short-term fluency into durable learning and transfer.

VII. CONCLUSIONS AND IMPLICATIONS

This review advances an AI–Learner Working-Memory (WM) co-regulation account that explains why educational AI alternately functions as a scaffold or a surrogate. The pivot points are assistance timing and output granularity: when help arrives after a learner attempt and remains short, local, and step-bound, AI reduces extraneous load, stages intrinsic complexity, and channels effort toward germane processing; when help precedes or replaces learner action—or arrives as global, verbose text—it displaces retrieval and inflates extraneous load, yielding near-term fluency but fragile, non-transferable knowledge. Treating AI as a designable scaffold tuned to the learner's WM profile reconciles mixed findings across tasks, levels, and disciplines and provides a principled basis for instruction, tools, and evaluation. Our incremental contribution is to recast AI–learner interaction as working-memory co-regulation by formalizing assistance timing and output granularity as design levers, and to pair this with a measurement agenda that combines tri-component cognitive-load instruments, process traces, and delayed transfer outcomes.

For instructional practice, courses should gate assistance by timing, require a brief pre-attempt to surface knowledge gaps, then supply minimal, local hints and close the loop with a delayed check (approximately 24–72 hours) to restabilize memory. Granularity should be constrained: stepwise or diagram-first explanations in STEM and outline-then-expand support in writing reduce search and split-attention costs while preserving the learner's generative work. Scope and novelty should be tightly controlled so that each prompt introduces only one new element, with clear signaling (e.g., numbering, headings) to guide processing. Scaffolds ought to fade quickly across isomorphic items to anticipate expertise-reversal effects for more advanced learners. Retrieval opportunities must be protected through low-stakes quizzes or short generative recalls immediately after AI use and again after a delay, and assessment should value process evidence such as pre-attempts and intermediate steps. Because divided attention erodes WM, learning environments should also minimize multitasking affordances and keep explanations as short as possible and as detailed as necessary. Cognitive barriers are required by default for products and policies, including verbosity limitations, salience for hallucination risk, and modes that prioritize incremental, local assistance Interfaces should make "scaffold not surrogate" a top priority for educators by promoting evidence of a pre-attempt before granting fuller assistance, scheduling delayed checks natively, and fading supports as competence signals improve. Analytics should render mechanisms visible by logging hint usage, dwell time, backtracks, step counts, and output length to diagnose extraneous-load inflation and inform instructional decisions. Transparency about model versions, settings, prompt templates, and guardrails—aligned with accessibility and academic-integrity requirements—supports safe and effective classroom adoption while prioritizing cognitive safety.

For research and evaluation, studies should explicitly manipulate and report the core design contrasts that the framework identifies: assistance timing (pre- versus post-attempt), granularity (stepwise/local versus global/long-form output), fading (fixed versus progressive removal of cues), and assessment placement (immediate versus delayed retrieval). A common battery will improve cumulation: pair tri-component load scales (intrinsic, extraneous, germane) and Paas mental-effort ratings with trace-based analytics such as hint usage, dwell time, and backtracks, and include delayed measures of recall and transfer at 24–72 hours. Mechanism-focused analyses should model mediation (extraneous decreases and germane increases) and moderation by prior knowledge, task complexity, grade level, and multitasking pressure, ideally using mixed-effects approaches that respect classroom clustering. Sharing representative prompt–output pairs, guardrails, and anonymized traces—and preregistering core contrasts where feasible—will further enable cumulative, mechanism-oriented science. Several boundary conditions and limitations merit caution. Large language model context windows are not human WM; analogies should remain heuristic rather than literal. Heterogeneity in tasks, prompts, and outcome measures still constrains cross-study comparability, and evidence for long-term transfer remains sparser than for short-term performance. These constraints reinforce the need for retrieval-preserving designs, careful control of verbosity and scope, and systematic inclusion of delayed outcomes in both research and practice. In sum, positioning AI as a scaffold rather than a surrogate offers a tractable route to convert short-term performance gains into durable learning and transfer. Time help after learner action, keep it short and local, fade it swiftly, and re-engage retrieval. Aligning instructional routines, product defaults, and evaluation methods around these principles organizes teaching, tooling, and evidence collection around a single mechanism: co-regulating WM so that assistance lightens avoidable load without displacing the cognitive work that makes knowledge last.

REFERENCES

1. Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International journal of educational technology in higher education*, *21*(1), 10.
2. Alanazi, M., Soh, B., Samra, H., & Li, A. (2025). The Influence of Artificial Intelligence Tools on Learning Outcomes in Computer Programming: A Systematic Review and Meta-Analysis. *Computers (2073-431X)*, *14*(5).
3. Andersen, M. S., & Makransky, G. (2021). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning*, *37*(1), 183-196.
4. Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K. (2023). A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review, 35*(1), Article 11. https://doi.org/10.1007/s10648-023-09745-1
5. Canonigo, A. M. (2024). Levering AI to enhance students' conceptual understanding and confidence in mathematics. *Journal of Computer Assisted Learning, 40*(6), 3215–3229. https://doi.org/10.1111/jcal.13065
6. Cardona, M. A., Rodríguez, R. J., & Ishmael, K**.** (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. U.S. Department of Education, Office of Educational Technology. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdfWang, J., Fan, W. The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanit Soc Sci Commun* 12, 621 (2025). https://doi.org/10.1057/s41599-025-04787-y
7. Centre for Education Statistics and Evaluation (CESE). 2017. *Cognitive load theory: Research that teachers really need to understand*. Sydney, New South Wales, Australia: CESE.
8. Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*(2), 233-246.
9. Choi, Y., & Lee, H. (2022). Psychometric Properties for Multidimensional Cognitive Load Scale in an E-Learning Environment. *International journal of environmental research and public health*, *19*(10), 5822. https://doi.org/10.3390/ijerph19105822
10. 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.
11. Dong, L., Tang, X., & Wang, X. (2025). Examining the effect of artificial intelligence in relation to students' academic achievement in classroom: A meta-analysis. *Computers and Education: Artificial Intelligence*, 100400.
12. Eğilmez, N. (2015). *The role of episodic memory in artificial intelligence* (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey
13. Endres, T., Carpenter, S., & Renkl, A. (2024). Constructive retrieval: Benefits for learning, motivation, and metacognitive monitoring. *Learning and Instruction*, 94, 101974.
14. Fedorets, V. M., Klochko, O. V., Tverdokhlib, I. A., & Sharyhin, O. A. (2024, October). Cognitive aspects of interaction in the "Human—Artificial Intelligence" system. In *Journal of Physics: Conference Series* (Vol. 2871, No. 1, p. 012023). IOP Publishing.
15. Fırat, M. (2013). Multitasking or continuous partial attention: A critical bottleneck for digital natives. *Turkish Online Journal of Distance Education*, *14*(1), 266-272.
16. Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 6.
17. Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging Cognitive Load Theory: The Role of Educational Neuroscience and Artificial Intelligence in Redefining Learning Efficacy. *Brain sciences*, *15*(2), 203. https://doi.org/10.3390/brainsci15020203
18. Gong, D., & Zhang, H. (2024). Self-attention limits working memory capacity of transformer-based models. *arXiv preprint arXiv:2409.10715*.
19. Gong, D., Wan, X., & Wang, D. (2024, March). Working memory capacity of ChatGPT: An empirical study. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 9, pp. 10048-10056).
20. Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly journal of experimental psychology (2006)*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060
21. Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research*, *33*(2), 678-696.
22. Hornby, R. (2025, July 8). *This is your brain on ChatGPT.* https://www.laptopmag.com/ai/chatgpt-mit-study-brain-activity-memory-impact

23. Huang, J. T., Sun, K., Wang, W., & Dredze, M. (2025). LLMs Do Not Have Human-Like Working Memory. *arXiv preprint arXiv:2505.10571*.
24. Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
25. Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *science*, *319*(5865), 966-968.
26. Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational psychology review*, *19*(3), 239-264.
27. Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458.
28. Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 61–77). Cambridge University Press.
29. Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. *arXiv preprint arXiv:2506.08872*.
30. Lademann, J., Henze, J., & Becker-Genschow, S. (2025). Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables. *Physical Review Physics Education Research*, *21*(1), 010147.
31. Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*(1), 65–100.
32. Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *NPJ science of learning*, *10*(1), 29. https://doi.org/10.1038/s41539-025-00320-7
33. Li, J., Yin, K., Wang, Y., Jiang, X., & Chen, D. (2025). Effectiveness of generative artificial intelligence-based teaching versus traditional teaching methods in medical education: a meta-analysis of randomized controlled trials. *BMC Medical Education*, 25(1), 1175.
34. Liebherr, M., Gößwein, E., Kannen, C., Babiker, A., Al-Shakhsi, S., Staab, V., ... & Montag, C. (2025). Working memory and the need for explainable AI–Scenarios from healthcare, social media and insurance. *Heliyon*, *11*(2).
35. Maddali, H. T., Dixon, E., Pradhan, A., & Lazar, A. (2022, April). Investigating the potential of artificial intelligence powered interfaces to support different types of memory for people with dementia. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1-7).
36. Martin, A. J., Collie, R. J., Kennett, R., Liu, D., Ginns, P., Sudimantara, L. B., ... & Rüschenpöhler, L. G. (2025). Integrating generative AI and load reduction instruction to individualize and optimize students' learning. *Learning and Individual Differences*, *121*, 102723.
37. Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693_eng
38. Nguyen, T. (2024). ChatGPT in medical education: a precursor for automation bias?. *JMIR Medical Education*, *10*, e50174.
39. Oakley, B., Johnston, M., Chen, K. Z., Jung, E., & Sejnowski, T. J. (2025). The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI. *arXiv preprint arXiv:2506.11015*.
40. OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing. https://doi.org/10.1787/c74f03de-en
41. Olawade, D. (2024). Memory systems and artificial intelligence: Linking human cognition and machine learning. *International Journal of School and Cognitive Psychology, 11*(S1), 1–2.
42. Ouwehand, K., Kroef, A. V. D., Wong, J., & Paas, F. (2021, September). Measuring cognitive load: Are there more valid alternatives to likert rating scales?. In *Frontiers in Education* (Vol. 6, p. 702616). Frontiers Media SA.
43. Pastötter, B., & Bäuml, K. H. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in psychology*, *5*, 286. https://doi.org/10.3389/fpsyg.2014.00286
44. Pastötter, B., & Frings, C. (2019). The forward testing effect is reliable and independent of learners' working memory capacity. *Journal of cognition*, *2*(1), 37.
45. Peng, J. L., & Yeh, S. L. (2025). Cognitive Offloading in Short-Term Memory Tasks: Trust Toward Tools as a Moderator. *International Journal of Human–Computer Interaction*, 1–12. https://doi.org/10.1080/10447318.2025.2474449

46. Qu, X., Sherwood, J., Liu, P., & Aleisa, N. (2025, April). Generative AI Tools in Higher Education: A Meta-Analysis of Cognitive Impact. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-9).
47. Raikov, A. (2021). *Cognitive semantics of artificial intelligence: a new perspective*. Singapore: Springer.
48. Renkl, A., & Atkinson, R. K. (2016). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. In *Cognitive Load Theory* (pp. 15-22). Routledge.
49. Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education, 70*(4), 293–315. https://doi.org/10.1080/00220970209599510
50. Reser, J. E. (2022). Artificial intelligence software structured to simulate human working memory, mental imagery, and mental continuity. *arXiv preprint arXiv:2204.05138*.
51. 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
52. Roediger , H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. *Journal of experimental psychology: applied*, *17*(4), 382.
53. Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, *15*(1), 20-27.
54. Romeo, G., & Conti, D. (2025). Exploring automation bias in human–AI collaboration: a review and implications for explainable AI. *AI & SOCIETY*, 1-20.
55. Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*(1), 12–19, 39.
56. Roy, P., Poet, H., Staunton, R., Aston, K., & Thomas, D. (2024, December 12). *ChatGPT in lesson preparation – A Teacher Choices trial*. National Foundation for Educational Research (NFER). https://www.nfer.ac.uk/publications/chatgpt-in-lesson-preparation-a-teacher-choices-trial/ NFER
57. Schlott, R. (2025, June 25). *Educators warn that AI shortcuts are already making kids lazy: "Critical thinking and attention spans have been demolished." New York Post*. https://nypost.com/2025/06/25/tech/educators-warn-that-ai-shortcuts-are-already-making-kids-lazy/
58. Shanmugasundaram, M., & Tamilarasu, A. (2023). The impact of digital technology, social media, and artificial intelligence on cognitive functions: a review. *Frontiers in Cognition*, *2*, 1203077.
59. 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.
60. Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of educational psychology*, *106*(2), 331.
61. Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.
62. Sweller, J., Ayres, P., & Kalyuga, S. (2011). The guidance fading effect. In *Cognitive load theory* (pp. 171-182). New York, NY: Springer New York.
63. The Times. (2025, June 25). *Using ChatGPT for work? It might make you more stupid*. https://www.thetimes.co.uk/article/using-chatgpt-for-work-it-might-make-you-more-stupid-dtvntprtk
64. Twabu, K. (2025). Enhancing the cognitive load theory and multimedia learning framework with AI insight. *Discover Education*, *4*(1), 160.
65. Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212-218.
66. Volzhenin, K., Changeux, J. P., & Dumas, G. (2022). Multilevel development of cognitive abilities in an artificial neural network. *Proceedings of the National Academy of Sciences*, *119*(39), e2201304119.
67. Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, *12*(1), 1-21.
68. Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49–64. https://doi.org/10.1080/00461520701756420
69. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0
70. Zheng, Q., Liu, H., Zhang, X., Yan, C., Cao, X., Gong, T., ... & Liu, J. (2025). Machine memory intelligence: Inspired by human memory mechanisms. *Engineering*.