# When Output Improves and Learning Does Not: The Cognitive Offloading and Engagement Framework for Generative AI Use

Amr Adel*, Mohammad AL-Rawi†, Tanmoy Debnath†, Manoranjan Paul‡
*School of Computing, Eastern Institute of Technology, Napier, New Zealand
†CSU Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia
‡School of Computing, Mathematics and Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia
avandenAdel@eit.ac.nz, {mal-rawi, tdebnath, mpaul}@csu.edu.au

***Abstract*—The rapid adoption of large language models (LLMs) has raised concerns about cognitive consequences, yet evidence remains fragmented across behavioural and neural studies. We ask whether generative AI use is associated with changes in human cognition. We conducted a structured search of a large multidisciplinary index (Semantic Scholar and OpenAlex via the Elicit semantic search engine), screening 499 retrieved records against pre-defined empirical criteria to yield 40 studies reporting behavioural, educational, or neurophysiological outcomes. Data were aligned across three domains: (1) task performance, (2) memory and learning retention, and (3) EEG connectivity. Synthesis reveals a consistent divergence: AI assistance reliably improves short-term performance (48% with basic interfaces, 127% with tutoring-style interfaces) but co-occurs with reduced directed EEG connectivity in semantic and monitoring networks (up to 55% reduction in longitudinal data), impaired memory encoding, and post-withdrawal performance declines (17%). To reconcile these conflicting findings, we propose the Cognitive Offloading and Engagement Framework, where cognitive engagement mediates the link between AI-assisted performance and a temporally accumulating construct termed cognitive debt. Effects appear to vary along an engagement–offloading continuum shaped by task structure and interface design, consistent with the pattern that scaffolded, tutoring-style interaction is more often associated with preserved engagement than unconstrained answer provision. The contribution is integrative and conceptual, consolidating dispersed evidence into a testable account with implications for human-AI interaction design. We note the evidence base remains small, largely correlational, and student-dominated.**

***Index Terms*—cognitive offloading, generative artificial intelligence, human–AI interaction, electroencephalography (EEG)**

## I. Introduction

The rapid diffusion of generative artificial intelligence systems, most notably large language models such as ChatGPT, has intensified debate over their cognitive consequences [1]–[4]. Public discourse frequently frames this debate in terms of intellectual decline, yet such claims often rely on anecdote, intuition, or short-term impressions rather than systematic evidence. The central question addressed in this paper is therefore not whether generative AI "feels" cognitively harmful, but whether its use is associated with measurable changes in human cognition that can be empirically observed, quantified, and compared across contexts. This question has become increasingly urgent given the speed and scale of adoption of large language models in education and knowledge work. Here goal structures determine whether ChatGPT is used for performance versus learning, with competitive performance goals driving memorisation rather than deep processing [5]. Generative AI tools are now routinely used for essay writing, programming, problem-solving, and creative ideation, frequently replacing or compressing cognitive processes that were previously effortful and time-consuming. Empirical studies consistently report short-term performance benefits during AI use, with a 2025 meta-analysis of experimental studies confirming ChatGPT enhances student learning performance when used as a scaffold [6]. Educational field experiments document performance improvements of 48% under basic generative AI interfaces and 127% under tutoring-oriented interfaces, while creative tasks show output gains of 3.7% to 9.0% during AI-assisted production [7]. This pattern is mirrored in workplace field data where generative AI increased issues resolved per hour by 15% on average and 35% for novices [8].

These results help explain the rapid uptake of such systems and their perceived productivity benefits. The divergence between immediate performance gains and longer-term cognitive outcomes is schematically summarised in Fig. 1, which highlights the temporal misalignment between efficiency signals and cognitive engagement. At the same time, an emerging body of evidence points to a tension between these immediate gains and longer-term cognitive outcomes. Behavioural studies report learning deficits following AI withdrawal, including 17% declines in grades once access is removed. Neurophysiological evidence further strengthens this concern. Longitudinal electroencephalography data show that users of large language models exhibit systematically weaker brain connectivity, with reductions of up to 55% in semantic and monitoring networks relative to unaided conditions, alongside impaired memory encoding and reduced ownership of written work [9].

The contribution of this paper lies in integrating behavioural, educational, and neurophysiological evidence into a unified analytical framework. Rather than treating performance gains and cognitive costs as isolated findings, we quantitatively align them to show how short-term efficiency improvements coexist

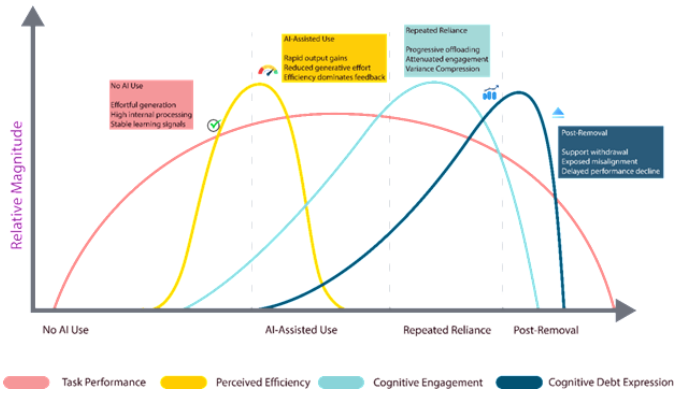



Fig. 1. Conceptual illustration of temporal divergence between task performance, perceived efficiency, cognitive engagement, and post-removal cognitive effects during generative AI use.

with measurable and persistent risks under conditions of unrestricted reliance. This integrative approach enables a clearer understanding of when generative AI functions as cognitive support and when it contributes to cognitive offloading with lasting consequences.

## II. Methodology

### A. Systematic Review Protocol

This study employs a structured systematic review methodology designed to identify empirical evidence linking generative AI use to measurable cognitive outcomes. A semantic search was conducted using the Elicit search engine, which indexes a large multidisciplinary corpus spanning Semantic Scholar and OpenAlex. The initial query retrieved 499 papers relevant to generative AI and cognitive functioning.

Screening was conducted using predefined inclusion criteria focused on empirical validity and cognitive relevance. Studies were retained only if they examined human participants using generative AI systems, measured cognitive outcomes such as learning, memory, critical thinking, creativity, or metacognition, and employed experimental, quasi-experimental, observational, longitudinal, or systematic review designs capable of assessing cognitive change. Opinion pieces, purely technical evaluations, and studies lacking behavioural or neurocognitive measurements were excluded. Applying these criteria resulted in a final corpus of 40 empirical studies, forming the evidentiary basis of the review.

### B. Data Extraction Strategy

For each included study, structured data extraction was performed across three domains. First, quantitative performance metrics were recorded, including percentage changes in task performance, grade outcomes, creativity scores, and effect sizes where reported. Second, neurophysiological measures were extracted when available, with particular attention to electroencephalography indicators such as directed neural connectivity, alpha and beta band engagement, and task-related network activation. Third, studies were classified according to temporal design, distinguishing short-term performance effects observed during AI use from longitudinal outcomes assessed after repeated exposure or AI withdrawal.

The longitudinal EEG experiment reported by Kosmyna et al. [9], involving 54 participants, four experimental sessions, and a four-month duration, was treated as a central empirical anchor for neurophysiological analysis, enabling alignment between behavioural outcomes and neural engagement patterns. Study quality was assessed narratively based on design features (sample size, controls, longitudinal follow-up).

### C. Integration and Analytical Approach

Rather than aggregating findings narratively, this review adopts a cross-study numerical alignment approach. This is a narrative rather than a statistical synthesis: outcome measures across the corpus (percentage grade change, standard-deviation effect sizes, EEG connectivity magnitude, p-values) are reported in incommensurable units and were not pooled quantitatively. Numerical alignment here refers to juxtaposing directionally consistent findings across domains, not to meta-analytic effect-size aggregation.

Quantitative results were compared across behavioural, educational, and neural domains to identify convergent patterns linking performance gains, cognitive engagement, and learning retention. Behavioural outcomes were explicitly mapped onto neurophysiological indicators, allowing performance improvements, connectivity reductions, and post-use declines to be analysed as interconnected components of a single cognitive trajectory. This integration strategy supports the development of a unified explanatory framework rather than a descriptive catalogue of effects.

## III. Literature Review: Quantitative Patterns Across Cognitive Domains

This literature review synthesises quantitative evidence across cognitive domains, focusing on effect magnitudes, statistical indicators, sample characteristics, and temporal dynamics. The review corpus comprises 40 sources screened from 499 papers identified through semantic search across over 138 million academic papers indexed in Elicit, including Semantic Scholar and OpenAlex.

### A. Study Base and Quantitative Coverage

The included evidence spans laboratory experiments, field experiments, quasi-experiments, observational longitudinal studies, and systematic reviews. Sample sizes range from small controlled studies, such as 16 students in classroom experiments, to large observational cohorts of 983 university students, and field experiments of approximately 1,000 high-school students. In creative-writing evaluation, one study used 293 writers and 600 evaluators. Several studies report participant age metrics, such as 117 university students (average age 22.61) and 61 college students (average age 21.56). The most detailed neurophysiological source in the corpus is Kosmyna et al. [9], which recruited 54 participants for Sessions 1 to 3, with 18 participants completing Session 4.

### B. Neural and Physiological Evidence of Cognitive Engagement

Across neurophysiological work, cognitive engagement scales downward with increasing external support. In the longitudinal EEG experiment, directed connectivity differences reached up to 55% reduced total directed transfer function magnitude in low-frequency semantic and monitoring networks among LLM users. Kosmyna et al. provide extensive quantitative methodological parameters that matter for interpretability and replication. EEG was collected using 32 electrodes at 500 samples per second with 24-bit resolution and measurement noise less than 1 $\mu$V RMS; filtering included 0.1 Hz and 100 Hz cut-offs and a 60 Hz notch filter [9]. Connectivity was computed across all directed pairs, 32 electrodes = 1024 total connections, with significance filtering at $p < 0.05$. Frequency bands were defined quantitatively, including Delta 0.1–4 Hz (subdivided into 0.2–0.83 Hz, 0.83–2.66 Hz, 2.66–4 Hz), Theta 4–8 Hz, Alpha 8–12 Hz (low 8–10 Hz, high 10–12 Hz), Beta 12–30 Hz (low 12–15 Hz, mid 15–18 Hz, high 18–30 Hz), and Gamma 30–100 Hz (low 30–44 Hz, high 44–100 Hz).

### C. Memory, Learning, and Retention

A central quantitative pattern is the divergence between gains during AI use and deficits after removal. In a field experiment of approximately 1,000 high-school mathematics students, access to GPT support increased performance by 48% using a basic interface and 127% using a tutoring interface, yet the basic-interface group later showed a 17% grade reduction relative to controls once access was removed [7]. Academic integrity mechanisms appear quantitatively linked to learning loss, with a 2025 meta-analysis of 41 studies reporting a moderate positive overall effect on academic achievement (g = 0.577) that disappears when AI is used as a substitute rather than scaffold [10]. In an introductory computer science context, a 25% plagiarism increase corresponded to a 10% reduction in exam scores, and students plagiarising every assignment scored 36 to 47% points lower on final exams [11]. In writing and memory encoding, Kosmyna et al. [9] report impaired quoting ability in LLM-assisted writers minutes after completion, consistent with shallow encoding and reduced ownership of produced text.

### D. Creativity and Divergent Thinking

Creativity outcomes show a quantitatively mixed profile, strongly moderated by task framing and user characteristics. Individual creative output during AI use increased by 3.7% to 9.0%, but collective novelty decreased as outputs became more similar [12]. Evidence also points to post-use costs. A field experiment reported subsequent innovation decreases of 0.6 to 0.72 SD after AI exposure [13]. A scaffolded ideation system produced statistically strong improvements, reported as $F(2, 8058.15) = 106.48$, $p < .0001$ [14]. Some collaborative creativity settings also report significance values, such as $p = 0.012$ for perceived creativity improvements under structured collaboration [15].

### E. Critical Thinking, Reasoning, and Problem-Solving

Critical-thinking results are sharply split between scaffolded and unstructured use, with survey evidence (n=666) showing AI tool use negatively predicts critical thinking and this effect is mediated by cognitive offloading [16]. A systematic review reports 62.5% of included studies finding positive critical-thinking effects, often concentrated at lower levels of Bloom's taxonomy [17]. This is consistent with a 2025 meta-analysis showing AI tools have a moderate positive effect on higher-order thinking skills when scaffolding is present [18].

In contrast, unstructured problem-solving shows striking numerical warning signs. In physics, 42% of AI-assisted solutions involved direct copy-paste behaviour, compared with 4% using search engines, and nearly half of solutions were mistakenly assumed correct [19]. In mathematics, ChatGPT assistance significantly decreased scores with $p = 0.002$ [15]. In programming implementation tasks, speed gains reached approximately three times faster completion, yet accompanied by reduced depth and uniformity of solutions [20].

### F. Summary Pattern

Across domains, quantitative evidence supports a consistent trajectory: measurable short-term performance gains coexist with reductions in cognitive engagement and measurable post-use deficits, particularly in learning retention and independent problem-solving. The most robust convergence appears where independent studies, despite differing populations and tasks, point in the same direction: field performance changes in one setting (48%, 127%, then 17% decline) and neurophysiological indicators of under-engagement in another (up to 55% reduced connectivity) are consistent with a common underlying pattern, though they were not measured within the same cohort and should not be read as a single causal chain [7], [9].

## IV. Neurophysiological Evidence and Mechanistic Interpretation

Rather than introducing new experimental data, the analysis draws primarily on electroencephalography findings, most notably those reported by Kosmyna et al. [9], and situates them within the broader quantitative patterns identified across the reviewed literature. Neurophysiological measures are treated as indicators of underlying cognitive processes that are not directly observable through performance metrics alone, offering one additional line of evidence in a sociotechnical context where causal claims are otherwise hard to establish. Neurophysiological evidence further supports this interpretation. As shown in Fig. 2, alpha-band dynamic directed transfer function analyses reveal systematically weaker functional connectivity during large language model use compared with brain-only conditions, consistent with reduced cognitive engagement.

### A. EEG Connectivity as an Index of Cognitive Engagement

Across studies employing EEG, cognitive engagement is operationalised through measures of functional and directed connectivity between cortical regions. Kosmyna et al. [9] report that participants using large language models during

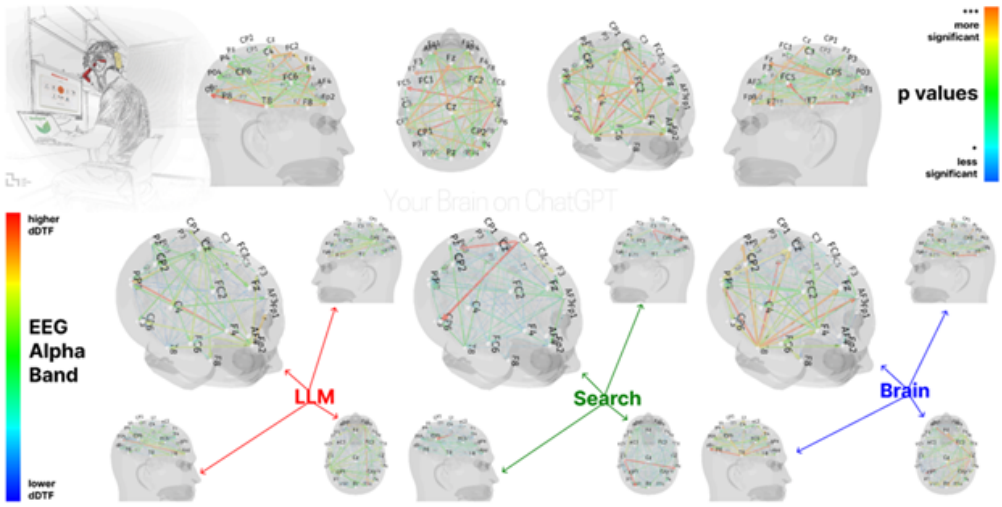



Fig. 2. Dynamic directed transfer function (dDTF) analysis of alpha-band EEG connectivity under language model, search engine, and brain-only conditions [9].

essay-writing tasks exhibited substantially weaker brain connectivity than participants working without AI assistance. Specifically, directed connectivity within low-frequency semantic and monitoring networks was reduced by up to 55%, as measured using directed transfer function magnitude. These networks are associated with integration of semantic content, self-monitoring, and executive coordination, functions that are central to effortful reasoning and learning.

The magnitude of this reduction persists across repeated sessions, though whether it exceeds normal within-task EEG variability has not been independently established. Importantly, connectivity reductions scaled with the degree of external assistance, indicating a systematic relationship between reliance on generative AI and neural under-engagement rather than a task-specific artefact.

### B. Frequency Band Analysis and Executive Control

Band-specific analyses provide further mechanistic insight. Alpha band activity, typically associated with attentional allocation and inhibitory control, was consistently attenuated in AI-assisted conditions, while beta band connectivity, linked to executive control and active maintenance of task goals, was also reduced [9]. This supports neuro-adaptive frameworks that align EEG P300 dynamics with generative AI interventions to preserve cognitive safety [21]. These patterns suggest that generative AI use dampens top-down control processes rather than merely redistributing cognitive effort.

Lower alpha and beta engagement implies reduced endogenous regulation of attention and action selection, consistent with a shift from generative processing to supervisory or verification-oriented behaviour. This interpretation aligns with behavioural findings of reduced mental effort, shallow processing, and diminished strategic thinking reported in parallel experimental studies.

### C. Cognitive Offloading and Neural Under-Engagement

Cognitive offloading [22] refers to the delegation of internal cognitive operations to external tools, thereby reducing the need for sustained internal processing, and recent 2026 evidence shows this offloading indirectly harms critical thinking via reduced cognitive self-efficacy [23]. Taken together, connectivity and frequency band findings support a cognitive offloading mechanism at the neural level. In the reviewed evidence, offloading is not inferred solely from self-report or behaviour but is directly observable as reduced coordination across neural networks responsible for semantic integration, monitoring, and executive control.

### D. Linking Neural Signatures to Behavioural Outcomes

The neurophysiological findings gain explanatory power when aligned with behavioural and educational outcomes. Reduced neural engagement co-occurs with impaired memory encoding, evidenced by a diminished ability to accurately quote one's own writing minutes after task completion. LLM-assisted production is likewise associated with lower perceived ownership of the resulting text [9]. Consistent with these findings, 2026 data confirm that cognitive offloading to generative AI predicts heightened concern over the erosion of cognitive ownership [24].

At the educational level, similar patterns are reflected in post-removal performance declines, including a 17% reduction in grades following AI withdrawal in a large field experiment and substantial exam score losses among students who repeatedly relied on AI-generated solutions. The convergence of neural under-engagement and behavioural deficits suggests that reduced connectivity is not epiphenomenal but functionally related to learning outcomes. Neural measures thus provide a bridge between short-term performance gains and longer-term cognitive costs, clarifying why improved outputs during AI use do not reliably translate into durable competence.

### E. Neurophysiology and Causal Plausibility in Sociotechnical Analysis

In debates about generative AI and cognition, causal claims are often contested because users self-select into particular usage patterns [25]–[28]. However, large-scale survey evidence from 319 knowledge workers shows higher confidence in AI is associated with self-reported reductions in cognitive effort and critical thinking across knowledge retrieval, comprehension and evaluation [29]. Neurophysiological evidence offers converging, though not yet independently replicated, support for causal interpretation by demonstrating within-participant changes in brain activity that track AI exposure over time. Because EEG measures capture real-time neural dynamics, they are less susceptible to demand characteristics, self-report bias, or grading artefacts.

When such systematic, quantitatively large, and temporally persistent reductions in connectivity are observed in association with generative AI use, they lend support to the claim that unrestricted reliance may alter core cognitive processes, though this remains to be independently confirmed [7], [11], [30], [31]. This does not imply that generative AI is inherently harmful, but it does demonstrate a biologically grounded mechanism through which sociotechnical design choices can shape cognitive engagement, learning, and autonomy.

## V. Conceptual Framework: Cognitive Offloading, Engagement, and Cognitive Debt

This section formalises a conceptual framework that integrates the quantitative behavioural patterns and neurophysiological signatures reviewed in earlier sections. The framework specifies how observed effects cohere mechanistically. Its purpose is to move beyond description and provide analytically precise constructs that explain why generative AI use can simultaneously enhance short-term performance and impair longer-term cognitive outcomes. Fig. 3 illustrates the Cognitive Offloading and Engagement Framework for Generative AI Use. The framework models interaction with generative AI as a staged cognitive process, beginning with information intake and attentional allocation, progressing through meaning construction and evaluative decision-making, and culminating in longer-term learning, adaptation, and cognitive consequences. Central to the framework is the distinction between task-level representations and propositional mental representations, and the treatment of cognitive offloading as a contingent outcome shaped by evaluative factors such as perceived sufficiency, effort, trust, and risk. The framework provides a unifying structure for analysing how short-term interaction decisions accumulate into enduring changes in user cognition.

### A. Cognitive Offloading as a Measurable Process

Cognitive offloading is defined here, following [22], not as a metaphorical reliance on tools, but as a measurable redistribution of cognitive labour from internally sustained neural processes to externally provided computational outputs. Within the reviewed evidence, offloading is operationalised through converging indicators. At the neural level, it manifests as systematic reductions in directed functional connectivity within semantic integration and monitoring networks, reaching up to 55% under conditions of generative AI assistance. At the behavioural level, it appears as substitution behaviours such as direct copy-paste strategies, increased plagiarism rates, and reduced revision depth.

These indicators scale with the degree of reliance on AI rather than appearing as binary effects, distinguishing dependent offloading, associated with greater AI control and lower intrinsic motivation, from autonomous offloading, where AI serves as scaffold [32]. This scaling property distinguishes cognitive offloading from simple task facilitation and supports its interpretation as a process that progressively displaces internally generated reasoning, monitoring, and consolidation.

### B. Cognitive Engagement as the Mediating Variable

Cognitive engagement functions as the central mediating variable linking AI use to learning outcomes. Engagement is understood as the sustained activation and coordination of neural systems responsible for attention regulation, executive control, semantic integration, and metacognitive monitoring. Neurophysiologically, engagement is indexed by the strength and distribution of alpha and beta band connectivity, which are consistently attenuated during AI-assisted tasks.

Within this framework, generative AI does not directly impair learning. Instead, it alters engagement profiles by reducing the need for endogenous control and generative processing. This reduction explains why task performance can improve during AI use while learning outcomes remain unchanged or deteriorate. Behavioural findings such as shallow encoding, impaired ability to quote one's own writing shortly after task completion, and reduced perceived ownership of outputs align with this interpretation, as each reflects diminished depth of processing rather than loss of capacity. This is a pattern confirmed by a 2025 meta-analysis showing mixed effects on behavioural, cognitive and emotional engagement depending on usage mode [33].

### C. Cognitive Debt and Temporal Accumulation

Following [9], we adopt cognitive debt to capture the temporal accumulation of learning deficits across behavioural and educational evidence. Cognitive debt refers to the accumulation of learning deficits that arise when repeated performance gains are achieved through reduced engagement rather than internal skill acquisition. The concept is grounded in two empirical regularities: persistence of neural under-engagement after AI withdrawal and measurable post-removal performance declines.

### D. Mapping Constructs to Empirical Indicators

The framework explicitly maps each construct to observable indicators. Cognitive offloading corresponds to reduced directed EEG connectivity, attenuation of alpha and beta band engagement, increased plagiarism, and solution substitution behaviours. Cognitive engagement is reflected in network coordination strength, revision depth, metacognitive monitoring

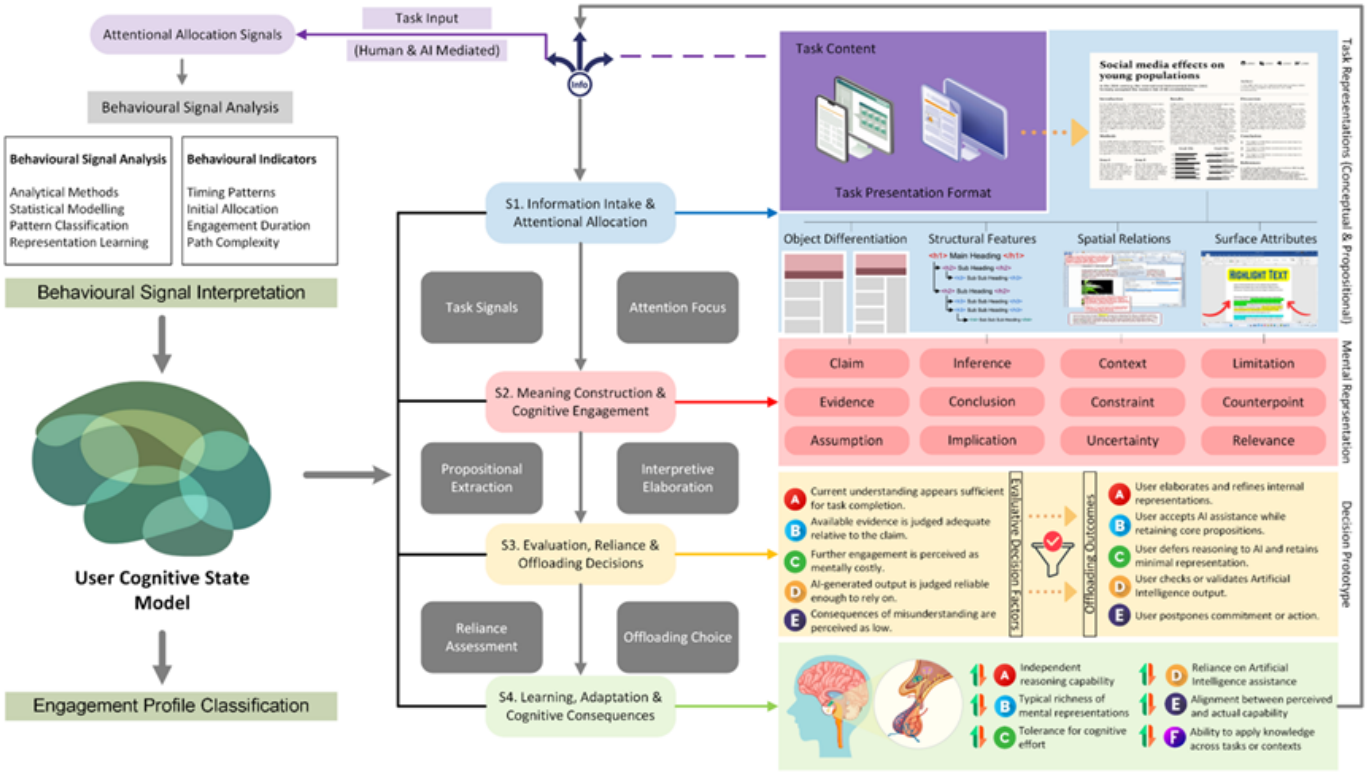



Fig. 3. The Cognitive Offloading and Engagement Framework for Generative AI Use.

behaviours, and accurate memory encoding. Cognitive debt is evidenced by the persistence of neural under-engagement, post-removal performance declines, and durable changes in creative and problem-solving behaviour.

This mapping ensures that the framework remains empirically anchored and testable. It also clarifies why short-term output quality is an insufficient proxy for cognitive impact, as performance metrics alone fail to capture engagement and its temporal consequences.

## VI. Discussion and Limitations

This section interprets the implications of the preceding synthesis without introducing new evidence. It addresses tensions in the literature, clarifies the scope of the claims advanced, and outlines the principal limitations that condition interpretation.

### A. Short-Term Gains Are Misleading Cognitive Indicators

A central implication of the reviewed evidence is that short-term performance gains achieved during generative AI use are systematically poor indicators of cognitive benefit, a pattern theorised as the augmentation trap where productivity gains coexist with cumulative costs of cognitive offloading [34]. Across domains, improvements in output quality, speed, or task completion coexist with markers of reduced cognitive engagement, shallow encoding, and diminished post-use performance.

Performance gains primarily reflect the efficiency of external computation, whereas learning depends on sustained engagement, internal generation, and metacognitive monitoring. When these processes are displaced, outputs may improve without corresponding skill acquisition. Treating short-term gains as evidence of cognitive enhancement therefore conflates productivity with competence and obscures longer-term risks. To explain why short-term performance gains coexist with post-removal decline, the study proposes a conceptual account of representational alignment and drift under repeated generative AI use. As illustrated in Fig. 4, representational alignment and drift explain why external anchoring during AI-assisted tasks compresses observable behavioural variability, while withdrawal exposes misalignment between surface performance and underlying cognitive state.

### B. Neurophysiology Matters for Sociotechnical Governance

Neurophysiological evidence plays a role in supporting, though not settling, causal interpretation within a sociotechnical debate that is otherwise dominated by behavioural proxies and self-report. Measures such as directed EEG connectivity and alpha and beta band engagement capture real-time cognitive dynamics that are not reducible to grading practices, motivation, or user preference. Because these measures track within-participant change over time, they help address concerns about self-selection and confounding. In governance contexts, this matters because policy decisions often hinge on whether observed outcomes reflect genuine cognitive change or superficial behavioural artefacts. Neurophysiology offers a biologically grounded, though currently thinly replicated bridge between system design and cognitive consequence, one

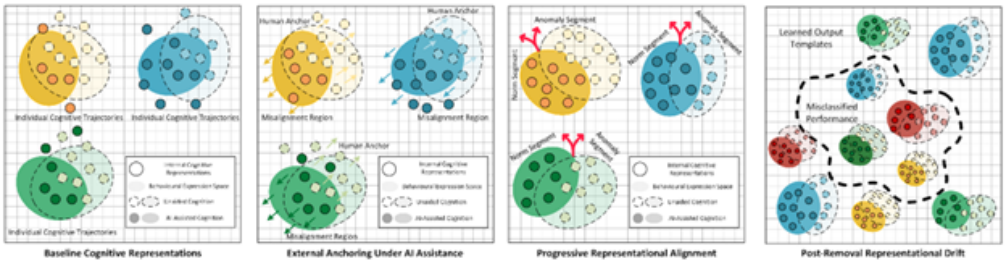



Fig. 4. Conceptual Model of Representational Alignment and Drift Under Generative AI Use.

that should inform rather than settle regulation and design guidance until corroborated by independent studies.

### C. Explaining Conflicting Findings in the Literature

The conceptual framework presented here resolves several apparent contradictions in the literature. Positive findings are concentrated in contexts where AI use preserves or enhances cognitive engagement through scaffolding, tutoring, or guided interaction. Recent experimental data show structured prompting significantly reduces cognitive offloading and enhances critical reasoning compared to unguided AI use [35].

Negative findings dominate where AI substitutes for generative effort through solution provision or copy-based use. By treating cognitive engagement as the mediating variable and cognitive debt as a temporal outcome, the framework explains why studies can report improved performance alongside impaired learning, and why effects differ by task type, interface design, and user ability. Conflicting results are thus not inconsistencies but reflections of different positions along the engagement–offloading spectrum, where heterogeneity in critical thinking emerges when thinking is outsourced without metacognitive engagement [36].

### D. Limits of Current Evidence

A further limitation concerns the evidentiary weight placed on a single neurophysiological anchor study. The EEG findings of [9], from which this paper adopts the term cognitive debt, derive from a sample of 54 participants across three sessions, falling to 18 in a fourth session, and the study has attracted published methodological criticism concerning its limited sample size, the reproducibility of its analyses, methodological aspects of the EEG analysis, inconsistencies in the reporting of results, and limited procedural transparency [37]. At the time of writing, the study remains a preprint rather than a peer-reviewed publication. Because this paper treats these findings as a central empirical anchor, the framework's neurophysiological grounding should be read as provisional pending independent replication, rather than as established mechanistic evidence. Relatedly, this review did not apply a formal risk-of-bias or study-quality instrument across the corpus; narrative quality judgements, rather than a standardised checklist, informed how heavily different studies were weighted in the synthesis.

Several limitations constrain the generality of current conclusions. First, the evidence base is heavily weighted toward student populations, particularly in higher education, limiting inference about other age groups and professional contexts. Second, task scope is narrow, with a strong emphasis on writing, programming, and problem-solving, leaving domains such as collaborative reasoning, long-term creative practice, and embodied tasks underexplored. Third, although longitudinal evidence exists, extended time horizons remain rare, and the durability of observed neural and behavioural changes beyond months is not yet established. Finally, neurophysiological studies, though methodologically strong, involve relatively small samples, which constrains fine-grained subgroup analysis.

## VII. Conclusion and Future Research Directions

This study set out to address a narrowly defined but consequential question: whether the use of generative AI is associated with measurable cognitive change, and if so, how such change should be interpreted. By synthesising quantitative behavioural outcomes with neurophysiological evidence, the paper moves beyond anecdotal claims of cognitive decline or uncritical assertions of cognitive enhancement. The central conclusion is not that generative AI is inherently cognitively harmful, but that its effects are conditional on how it is used and, critically, on whether its use preserves or displaces cognitive engagement.

Across domains, the evidence shows a consistent pattern. Generative AI reliably improves short-term task performance and output quality, yet these gains frequently coexist with reduced cognitive engagement, weaker memory encoding, and diminished independent performance once AI support is withdrawn. Neurophysiological findings provide a mechanistic anchor for this pattern, demonstrating systematic reductions in neural connectivity associated with attention, executive control, and semantic monitoring. When aligned with behavioural indicators such as shallow encoding, plagiarism, and post-removal decline, these findings support a coherent explanation grounded in cognitive offloading and the accumulation of cognitive debt.

The contribution of this paper lies in formalising that explanation. By defining cognitive offloading as a measurable process, identifying engagement as the mediating variable, and extending [9]'s construct of cognitive debt into a temporally accumulating outcome across domains, the framework aims to

reconcile conflicting findings in the literature while remaining explicit about the provisional strength of its neurophysiological grounding. It clarifies why productivity gains cannot be treated as proxies for learning and why sociotechnical design choices matter for cognitive outcomes.

Future research should extend this work in several constrained and testable directions. Longer-term longitudinal studies are needed to determine the durability and reversibility of observed neural and behavioural changes. Broader populations and task domains should be examined to assess generalisability beyond educational writing and problem-solving contexts. Finally, design-based research evaluating scaffolded interaction patterns can help establish conditions under which generative AI supports learning without inducing disengagement.

## REFERENCES


[1] L. G. Conti and P. Seele, "Reflexivity and positionality statements: A way to tackle "second-level arbitrariness" bias in AI?" *Humanities and Social Sciences Communications*, vol. 12, no. 1, p. 1930, 2025.

[2] X. Hao, J. Xu, and Y. Wang, "How generative AI shapes user perceived value and adoption intention in digital museum experiences," *npj Heritage Science*, vol. 13, no. 1, p. 608, 2025.

[3] H. Lin, G. Czarnek, B. Lewis, J. P. White, A. J. Berinsky, T. Costello, G. Pennycook, and D. G. Rand, "Persuading voters using human–artificial intelligence dialogues," *Nature*, vol. 648, no. 8093, pp. 394–401, 2025.

[4] J. Zhang, W. Song, and Y. Liu, "Cognitive bias in generative AI influences religious education," *Scientific Reports*, vol. 15, no. 1, p. 15720, 2025.

[5] L. Schmidt, N. Obergassel, and J. Roelle, "Aiming high: Do goal structures matter in learning with chatgpt?" *Applied Cognitive Psychology*, vol. 39, no. 6, p. e70148, 2025.

[6] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does chatgpt enhance student learning? a systematic review and meta-analysis of experimental studies," *Computers & Education*, vol. 227, p. 105224, 2025.

[7] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI without guardrails can harm learning: Evidence from high school mathematics," *Proceedings of the National Academy of Sciences*, vol. 122, no. 26, p. e2422633122, 2025.

[8] E. Brynjolfsson, D. Li, and L. R. Raymond, "Generative ai at work," *The Quarterly Journal of Economics*, vol. 140, no. 2, pp. 889–942, 2025.

[9] N. Kosmyna, E. Hauptmann, Y. T. Yuan, J. Situ, X.-H. Liao, A. V. Beresnitzky, I. Braunstein, and P. Maes, "Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task," 2025.

[10] Z. Liu, H. Zuo, and Y. Lu, "The impact of chatgpt on students' academic achievement: A meta-analysis," *Journal of Computer Assisted Learning*, vol. 41, p. e70096, 2025.

[11] B. Chen, C. M. Lewis, M. West, and C. Zilles, "Plagiarism in the age of generative AI: Cheating method change and learning loss in an intro to CS course," in *Proceedings of the Eleventh ACM Conference on Learning @ Scale*. ACM, 2024, pp. 75–85.

[12] A. R. Doshi and O. P. Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," *Science Advances*, vol. 10, no. 28, p. eadn5290, 2024.

[13] M. Hassan, E. Amin, S. Mansour, and Z. Kelani, "Is ChatGPT detrimental to innovation? A field experiment among university students," 2024.

[14] J. L. Heyman, S. R. Rick, G. Giacomelli, H. Wen, R. Laubacher, N. Taubenslag, M. Knicker, Y. Jeddi, P. Ragupathy, J. Curhan, and T. Malone, "Supermind Ideator: How scaffolding human–AI collaboration can increase creativity," in *Proceedings of the ACM Collective Intelligence Conference*. ACM, 2024, pp. 18–28.

[15] D. G. Kim and A. Moon, "From helping hand to stumbling block: The ChatGPT paradox in competency experiment," *Applied Economics Letters*, vol. 32, no. 18, pp. 2669–2673, 2025.

[16] M. Gerlich, "Ai tools in society: Impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, p. 6, 2025.

[17] J. Sardi, Darmansyah, O. Candra, D. F. Yuliana, Habibullah, D. T. P. Yanto, and F. Eliza, "How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review," *International Journal of Engineering Pedagogy*, vol. 15, no. 1, p. 94, 2025.

[18] M. Ilgun Dibek, M. Sahin Kursad, and T. Erdogan, "Influence of artificial intelligence tools on higher order thinking skills: a meta-analysis," *Interactive Learning Environments*, vol. 33, no. 3, pp. 2216–2238, 2025.

[19] L. Krupp, S. Steinert, M. Kiefer-Emmanouilidis, K. E. Avila, P. Lukowicz, J. Kuhn, S. Küchemann, and J. Karolus, "Unreflected acceptance—Investigating the negative consequences of ChatGPT-assisted problem solving in physics education," 2023.

[20] N. Baláž, J. Porubän, M. Horváth, and T. Kormaník, "Using ChatGPT during implementation of programs in education," in *5th International Computer Programming Education Conference (ICPEC 2024)*, ser. OASIcs, A. L. Santos and M. Pinto-Albuquerque, Eds., vol. 122. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024, pp. 18:1–18:9.

[21] L. Yan *et al.*, "Nested learning in higher education: Integrating generative ai, neuroimaging, and multimodal deep learning for a sustainable and innovative ecosystem," *Sustainability*, vol. 18, no. 2, p. 656, 2026.

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

[23] J. Wang, "Cognitive offloading through digital tools and its relationship with critical thinking, task persistence, and learning depth," *Frontiers in Psychology*, vol. 17, p. 1781101, 2026.

[24] J. Su, Z. Wang, Z. Qiang, and Z. Wang, "Is it mine or not mine? cognitive offloading to generative ai and concern over the erosion of cognitive ownership," *Journal of Intelligence*, vol. 14, no. 8, p. 182, 2026.

[25] C. Gao, X. Lan, N. Li, Y. Yuan, J. Ding, Z. Zhou, F. Xu, and Y. Li, "Large language models empowered agent-based modeling and simulation: A survey and perspectives," *Humanities and Social Sciences Communications*, vol. 11, no. 1, p. 1259, 2024.

[26] T. Niu, T. Liu, Y. T. Luo, P. C.-I. Pang, S. Huang, and A. Xiang, "Decoding student cognitive abilities: A comparative study of explainable AI algorithms in educational data mining," *Scientific Reports*, vol. 15, no. 1, p. 26862, 2025.

[27] W. Wang and Y. Lu, "Survey on Chinese users' acceptance of AI assistants: Expanding technology acceptance model," *Scientific Reports*, vol. 15, no. 1, p. 33535, 2025.

[28] Y. Zhai and B. Nezakatgoo, "Evaluating AI-powered applications for enhancing undergraduate students' metacognitive strategies, self-determined motivation, and social learning in English language education," *Scientific Reports*, vol. 15, no. 1, p. 35199, 2025.

[29] H.-P. Lee, A. Sarkar, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, and N. Wilson, "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 CHI Conference on Human Factors in Computing Systems*, 2025.

[30] Y. Fan, L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, Y. Shen, X. Li, and D. Gašević, "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance," *British Journal of Educational Technology*, vol. 56, no. 2, pp. 489–530, 2025.

[31] J. Tian and R. Zhang, "Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy," *Acta Psychologica*, vol. 260, p. 105725, 2025.

[32] Q. Zhu *et al.*, "Not all cognitive offloading is equal: distinguishing dependent and autonomous offloading to generative ai," *Frontiers in Psychology*, 2026.

[33] Y. M. E. Heung and T. K. F. Chiu, "How chatgpt impacts student engagement from a systematic review and meta-analysis study," *Computers and Education: Artificial Intelligence*, vol. 8, p. 100361, 2025.

[34] R. Sharma and L. Chen, "The augmentation trap: Ai productivity and the cost of cognitive offloading," 2026.

[35] M. Gerlich, "From offloading to engagement: An experimental study on structured prompting and critical reasoning with generative ai," *Data*, vol. 10, no. 11, p. 172, 2025.

[36] W. Zhang, M. Li, and Y. Chen, "When thinking is outsourced: Cognitive offloading and the heterogeneity of critical thinking among chinese university students using generative artificial intelligence," *Journal of Intelligence*, vol. 14, no. 7, p. 116, 2026.

[37] M. Stankovic, E. Hirche, S. Kollatzsch, and J. N. Doetsch, "Comment on: Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing tasks," 2026.