# Deep Research Prompt: AI and Cognitive Offloading in Higher Education Conduct comprehensive research on the use of AI tools, particularly large language models (LLMs), in higher education, with emphasis on their cognitive and learning effects. ## Research Focus Areas ### Core Investigation - **Cognitive offloading effects**: How students' reliance on AI tools affects learning outcomes, retention, and skill development - **Productivity vs. learning effectiveness**: The distinction between using AI to complete tasks faster versus using AI to understand and learn more deeply - **Cognitive load theory**: Examine how AI tools modify intrinsic, extraneous, and germane cognitive load in educational contexts ### Theoretical Framework Draw primarily from: - Cognitive psychology research on learning, memory, and skill acquisition - Cognitive load theory (Sweller, Paas, van Merriënboer) - Desirable difficulties and effortful learning (Bjork) - Distributed cognition and extended mind theory - Research on calculator/GPS/spell-checker effects as historical analogues ### Key Questions to Address 1. When does AI reduce beneficial cognitive effort ("desirable difficulty") versus harmful cognitive overload? 2. What evidence exists for skill atrophy or development when students use AI assistance? 3. How do different types of AI engagement (generation, editing, critique, explanation) affect learning? 4. What distinguishes surface-level offloading from strategic cognitive partnerships? 5. Are there differential effects across disciplines, task types, or student expertise levels? ## Research Parameters ### Sources to Prioritize - Peer-reviewed empirical studies on AI/LLM use in education (2020-present) - Cognitive psychology foundations (seminal works + recent applications to AI) - Educational psychology research on learning strategies and metacognition - Critical analyses from learning scientists and cognitive researchers - Reports from educational institutions implementing AI tools ### Perspective - Maintain analytical rigor and evidence-based reasoning - Adopt a **moderately critical stance**: acknowledge potential benefits while rigorously examining risks - Highlight gaps between productivity gains and learning gains - Question techno-optimist assumptions about AI in education - Consider both immediate task performance and long-term learning outcomes ### Structure Requirements Organize findings into: 1. **Executive summary** with key findings and implications 2. **Cognitive load analysis**: How AI affects different load types 3. **Productivity-learning tension**: Evidence for when these diverge 4. **Empirical evidence**: Studies on learning outcomes with AI assistance 5. **Theoretical concerns**: Cognitive psychology perspectives on potential risks 6. **Nuanced applications**: When/how AI might genuinely support learning 7. **Research gaps** and future investigation needs 8. **Practical implications** for educators and institutions ###

retention, and metacognitive skills, while structured pedagogical scaffolding that reserves higher-order cognitive tasks for human effort can mitigate these risks and support genuine learning.

## Abstract

This systematic review synthesizes findings from 25 empirical studies examining how AI tools, particularly large language models, affect cognitive processes and learning outcomes in higher education. The evidence reveals a fundamental tension between productivity and learning: AI tools reliably improve immediate task performance across contexts, but these gains do not automatically translate to deeper learning or skill development. A meta-analysis found that generative AI significantly enhances lower-order cognitive outcomes (understanding, applying) with an overall effect size of 1.34, yet impact on higher-order skills (analyzing, evaluating, creating) was minimal . Studies employing rigorous measures document concerning patterns of cognitive cost—high AI dependency correlates with 17.3 percentage points lower critical thinking scores and 22% worse memory retention , while neuroimaging evidence reveals significantly weaker brain connectivity in LLM users compared to those working without AI assistance . Longitudinal findings show these costs accumulate over time, with students converging toward passive engagement patterns and experiencing measurable declines in metacognitive skills .

However, the evidence also demonstrates that instructional design critically moderates these effects. Structured, guided AI use with deliberate pedagogical scaffolding can support learning when implementation reserves higher-order cognitive tasks for human effort while delegating lower-order tasks to AI . The divergence in findings across studies maps systematically onto implementation characteristics: positive outcomes occur with instructor guidance and structured workflows, while negative outcomes emerge from free, unstructured, or high-intensity use . Critical thinking disposition also moderates outcomes, with students high in critical thinking showing reflective and cautious AI use while those with high trust but low critical thinking demonstrate thoughtless use patterns . Significant methodological limitations constrain current conclusions, including reliance on self-report measures, limited longitudinal data, and a pervasive gap between productivity metrics commonly assessed and learning outcomes that matter educationally. The evidence supports neither wholesale adoption nor rejection of AI in education, but rather indicates that cognitive offloading effects depend fundamentally on how AI tools are integrated into pedagogical practice.

## Paper search

We performed a semantic search using the query "# Deep Research Prompt: AI and Cognitive Offloading in Higher Education

Conduct comprehensive research on the use of AI tools, particularly large language models (LLMs), in higher education, with emphasis on their cognitive and learning effects.

## Research Focus Areas

### Core Investigation

- **Cognitive offloading effects**: How students' reliance on AI tools affects learning outcomes, retention, and skill development
- **Productivity vs. learning effectiveness**: The distinction between using AI to complete tasks faster versus using AI to understand and learn more deeply
- **Cognitive load theory**: Examine how AI tools modify intrinsic, extraneous, and germane cognitive load in educational contexts

### Theoretical Framework

Draw primarily from:

- Cognitive psychology research on learning, memory, and skill acquisition
- Cognitive load theory (Sweller, Paas, van Merriënboer)
- Desirable difficulties and effortful learning (Bjork)
- Distributed cognition and extended mind theory
- Research on calculator/GPS/spell-checker effects as historical analogues

### Key Questions to Address

1. When does AI reduce beneficial cognitive effort ("desirable difficulty") versus harmful cognitive overload?
2. What evidence exists for skill atrophy or development when students use AI assistance?
3. How do different types of AI engagement (generation, editing, critique, explanation) affect learning?
4. What distinguishes surface-level offloading from strategic cognitive partnerships?
5. Are there differential effects across disciplines, task types, or student expertise levels?

## Research Parameters

### Sources to Prioritize

- Peer-reviewed empirical studies on AI/LLM use in education (2020-present)
- Cognitive psychology foundations (seminal works + recent applications to AI)
- Educational psychology research on learning strategies and metacognition
- Critical analyses from learning scientists and cognitive researchers
- Reports from educational institutions implementing AI tools

### Perspective

- Maintain analytical rigor and evidence-based reasoning
- Adopt a **moderately critical stance**: acknowledge potential benefits while rigorously examining risks
- Highlight gaps between productivity gains and learning gains
- Question techno-optimist assumptions about AI in education
- Consider both immediate task performance and long-term learning outcomes

### Structure Requirements

Organize findings into:

1. **Executive summary** with key findings and implications
2. **Cognitive load analysis**: How AI affects different load types
3. **Productivity-learning tension**: Evidence for when these diverge
4. **Empirical evidence**: Studies on learning outcomes with AI assistance
5. **Theoretical concerns**: Cognitive psychology perspectives on potential risks
6. **Nuanced applications**: When/how AI might genuinely support learning
7. **Research gaps** and future investigation needs
8. **Practical implications** for educators and institutions

**Critical Analysis Elements**

- Distinguish correlation from causation in existing studies
- Note methodological limitations in current research
- Identify conflicts between short-term metrics and long-term learning
- Examine whether AI tools align with or undermine cognitive principles of effective learning
- Consider equity implications of cognitive offloading

## Output Specifications

- Comprehensive report (8,000-12,000 words)
- Extensive citations with full bibliographic information
- Evidence-based conclusions with appropriate epistemic humility
- Actionable insights for higher education contexts
- Balanced treatment that doesn't dismiss benefits but foregrounds learning-oriented concerns

Focus on synthesizing insights that bridge cognitive science theory and educational practice, with particular attention to unintended consequences and understudied risks of AI-mediated learning." across over 138 million academic papers from the Elicit search engine, which includes all of Semantic Scholar and OpenAlex.

We retrieved the 500 papers most relevant to the query.

## Screening

We screened in sources based on their abstracts that met these criteria:

- **Higher Education Population**: Does this study involve undergraduate or graduate students in college, university, or equivalent post-secondary educational settings?
- **AI Tool Educational Intervention**: Does this study examine the use of artificial intelligence tools (such as large language models, chatbots, writing assistants, or tutoring systems) in academic or learning contexts?
- **Cognitive or Learning Outcomes**: Does this study assess at least one cognitive or learning outcome (such as learning performance, retention, skill development, cognitive load, metacognition, or academic achievement) rather than only measuring satisfaction, attitudes, or usage patterns?
- **Empirical Evidence**: Does this study present original empirical data (quantitative, qualitative, or mixed methods) or systematic synthesis of empirical evidence, rather than being an opinion piece, editorial, or purely theoretical paper without empirical support?
- **Educational Task Context**: Does this study examine AI use in authentic educational tasks (such as writing, problem-solving, research, studying, or skill acquisition) rather than focusing on administrative purposes, institutional management, or non-academic applications?
- **Publication Quality**: Is this study published in a peer-reviewed journal, peer-reviewed conference proceedings, or does it represent rigorous institutional research?
- **Adequate Sample and Scope**: Does this study include an adequate sample size (at least 10 participants if it's a case study) or, if smaller, does it offer unique methodological innovations or document rare phenomena?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

## Data extraction

We asked a large language model to extract each data column below from each paper. We gave the model the extraction instructions shown below for each column.

- **AI Tool Type**:

  Extract details about the AI tools and interaction types studied, including:

- Specific AI system(s) used (e.g., ChatGPT, GPT-4, generic LLM)
- Type of AI assistance provided (e.g., text generation, editing, explanation, debugging, brainstorming)
- Level of AI involvement (e.g., complete task completion, partial assistance, guidance only)
- Student control over AI interaction (e.g., free use, structured prompts, instructor-guided)
- Comparison conditions (e.g., no AI, search engines, traditional tools)

- **Cognitive Measurement**:

  Extract all methods used to assess cognitive effects, including:

- Cognitive load measures (e.g., EEG, self-report, task performance indicators)
- Learning outcome assessments (e.g., knowledge tests, skill demonstrations, retention measures)
- Cognitive skill level targeted (e.g., Bloom's taxonomy levels, higher-order vs lower-order thinking)
- Neural measures if used (e.g., brain connectivity, activation patterns)
- Metacognitive assessments (e.g., self-regulation, awareness of learning process)
- Timeline of measurement (immediate, delayed, longitudinal)

- **Learning Context**:

  Document the educational context and task characteristics:

- Academic discipline/subject area
- Task complexity and type (e.g., essay writing, programming, problem-solving, reading comprehension)
- Student population (e.g., undergraduate/graduate, expertise level, demographics)
- Instructional approach (e.g., guided AI use, unstructured access, pedagogical scaffolding)
- Duration and intensity of AI exposure
- Assessment/grading implications of AI use

- **Productivity vs Learning**:

  Extract evidence distinguishing between productivity and learning effects:

- Task completion metrics (e.g., speed, quantity, immediate output quality)
- Learning effectiveness measures (e.g., understanding, skill transfer, retention)
- Explicit comparisons between efficiency gains and learning gains
- Evidence of trade-offs between quick task completion and deeper learning
- Long-term vs short-term performance differences
- Student perceptions of productivity vs learning benefits

- **Cognitive Load Effects**:

  Document findings related to cognitive load theory:

- Effects on intrinsic cognitive load (task difficulty)
- Effects on extraneous cognitive load (poor instruction/interface)

- Effects on germane cognitive load (productive mental effort)
- Evidence of cognitive effort reduction or redistribution
- Impact on 'desirable difficulties' and effortful learning
- Cognitive engagement patterns over time
- Brain activity/connectivity changes if measured

- **Skill Development Impact**:

  Extract evidence about effects on skill acquisition and retention:

- Skill development outcomes (improvement, maintenance, decline)
- Evidence of skill atrophy or dependency on AI tools
- Transfer of learning to non-AI contexts
- Development of AI-independent capabilities
- Changes in problem-solving strategies or approaches
- Long-term competency effects
- Comparison of skills in AI vs non-AI conditions

- **Engagement Patterns**:

  Document how students actually used AI tools and engagement changes:

- Types of prompts or interactions used
- Evolution of AI usage patterns over time
- Active vs passive engagement behaviors
- Strategic vs surface-level use of AI
- Self-regulation and metacognitive behaviors with AI
- Over-reliance or dependency indicators
- Shifts in cognitive engagement over repeated exposure

- **Key Findings**:

  Extract the main results and conclusions about cognitive/learning effects:

- Primary cognitive outcomes (positive, negative, neutral)
- Effect sizes and statistical significance where reported
- Differential effects by task type, student level, or context
- Identified mechanisms explaining the effects
- Unexpected or counterintuitive findings
- Distinctions between immediate and long-term effects
- Author conclusions about educational implications
- Identified risks or benefits for learning

- **Study Limitations**:

  Document methodological constraints and research gaps:

- Sample size and duration limitations
- Measurement challenges or limitations
- Confounding variables or alternative explanations
- Generalizability constraints
- Author-acknowledged limitations

- Missing cognitive measures or comparisons
- Gaps between productivity metrics and learning assessment
- Need for longer-term or different types of studies

## Characteristics of Included Studies

This systematic review synthesizes findings from 25 empirical studies and reviews examining the cognitive and learning effects of AI tools in higher education contexts. The studies span multiple disciplines, methodological approaches, and geographic regions, providing a comprehensive view of how large language models and generative AI affect student cognition and learning.

| Study | Full text retrieved? | Study Type | AI Tools Examined | Primary Domain | Sample Size |
|---|---|---|---|---|---|
| Hareem Arif et al., 2025 | Yes | Structured review | ChatGPT, Grammarly | EFL Writing | Multiple studies synthesized |
| Nataliya Kosmyna et al., 2025 | Yes | Experimental with EEG | ChatGPT/LLMs | Essay writing | 54 participants |
| Hui Hong et al., 2025 | No | Quasi-experimental | Generative AI | English essay writing | 240 students |
| Xiaodong Qu et al., 2025 | Yes | Meta-analysis | ChatGPT | Multiple disciplines | 17 studies synthesized |
| Dr. Atiya Rohilla et al., 2025 | Yes | Mixed methods | ChatGPT, Grammarly, Carnegie Learning | STEM, humanities, social sciences | n=300 survey, n=45 interviews |
| Majeed Kazemitabaar et al., 2024 | Yes | Classroom deployment | CodeAid (GPT-3/3.5 based) | Programming | 700 students, 8,000 interactions |
| Hui Hong et al., 2025a | No | Quantitative | Generative AI | English essay writing | 240 students |
| Yue Fu et al., 2025 | Yes | Observational | ChatGPT, Gemini, Perplexity, Claude | Academic reading | 124 sessions |
| Alexandru Dinu et al., 2025 | Yes | Mixed methods quasi-experimental | ChatGPT, ChatGPT 4o | Engineering | 60 students |
| Utkarsh Arora et al., 2024 | Yes | Observational | ChatGPT (GPT-3.5/4), GitHub Copilot | Distributed Systems | 411 students |
| M. Mutanga et al., 2025 | No | Mixed methods | LLMs | Programming | 140 students, 842 prompts |

| Study | Full text retrieved? | Study Type | AI Tools Examined | Primary Domain | Sample Size |
|---|---|---|---|---|---|
| Joseph Crawford et al., 2024 | Yes | Survey with SEM | ChatGPT | General academic | 387 students |
| Sanka Rasnayaka et al., 2024 | Yes | Semester-long deployment | GitHub Copilot, ChatGPT | Software engineering | 214 students |
| Min Zhang et al., 2025 | Yes | Experimental | GPT-4 | Academic writing | 246 students |
| Yaoying Han et al., 2025 | Yes | Mixed methods longitudinal | ChatGPT, Grammarly | Academic writing | 300 students, 45 educators |
| Sidra Zaheer et al., 2025 | No | Systematic review | Grammarly, Criterion, Quillbot, ChatGPT | English academic writing | 46 articles reviewed |
| Samuel Boguslawski et al., 2024 | Yes | Qualitative interviews | LLMs | Programming | 12 students, 6 faculty |
| Hong Zhang et al., 2025 | No | Qualitative observational | DeepSeek/ChatGPT | EFL writing | 87 students |
| M. Maphalala et al., 2025 | Yes | Systematic review | General AI tools | Higher education broadly | 87 studies reviewed |
| Viktoriia Pryma et al., 2025 | Yes | Mixed methods | Grammarly, QuillBot | English for Academic Purposes | 60 students |
| A. Bernik et al., 2025 | Yes | Experimental | ChatGPT-4, Claude 3, Gemini 1.5 Pro | Programming assessment | 315 assignments |
| Hardiyanti Pratiwi et al., 2025 | No | Phenomenological | ChatGPT, Deepl, Zotero, Scite AI | Doctoral research | 81 participants |
| Ruiwei Xiao et al., 2025 | Yes | Intervention study | GPT-4o | CS education | 22 students + 36 instructors |
| Chenyu Hou et al., 2025 | Yes | Survey with path analysis | Generative AI | Digital literacy | 808 students |
| Zulaikha Zulaikha et al., 2025 | No | Quasi-experimental | ChatGPT, Grammarly, Quilbolt | Arabic Language Education | 90 students |

The included studies represent diverse methodological approaches, with quasi-experimental designs being most common, supplemented by observational studies, systematic reviews, and one study employing neuroimaging (EEG). The predominant domains examined were writing (both L1 and L2 contexts) and programming education. Sample sizes ranged from 22 to 808 participants in primary studies, with meta-analyses and systematic reviews synthesizing larger bodies of evidence. Notably, only one study (Kosmyna et al.) employed neural measures to directly assess cognitive engagement , representing a significant methodological gap in the literature.

## Cognitive Load Effects

### Theoretical Framework and Findings

The studies provide substantial evidence regarding how AI tools modify the three components of cognitive load identified by Sweller's cognitive load theory. The effects on intrinsic, extraneous, and germane cognitive load appear to be complex and often contradictory, depending heavily on implementation context and pedagogical scaffolding.

| Study | Effect on Intrinsic Load | Effect on Extraneous Load | Effect on Germane Load |
|---|---|---|---|
| Hareem Arif et al., 2025 | Reduced through immediate feedback | Potentially increased through dependency | Enhanced under scaffolding, diminished if uncritical |
| Nataliya Kosmyna et al., 2025 | Reduced due to ease of access | Reduced by streamlining information | Compromised due to reduced active integration |
| Hui Hong et al., 2025 | Not measured | Reduced by delegating lower-order tasks | Enhanced focus on higher-order thinking |
| Xiaodong Qu et al., 2025 | Reduced for lower-order skills | Reduced through guided prompts | Limited impact on higher-order skills |
| Dr. Atiya Rohilla et al., 2025 | May cause cognitive underload | Reduced by automating lower-order activities | Decreased due to reduced effortful processing |
| Min Zhang et al., 2025 | Not explicitly measured | Guided use reduces extraneous load | Enhanced through structured guidance |
| Yaoying Han et al., 2025 | Reduced during drafting (28-35% improvement) | Unrestricted access creates "cognitive crutches" | Heavy reliance leads to 15% decline in metacognitive skills |

The meta-analysis by Qu et al. provides critical insight into the differential effects across Bloom's taxonomy levels. GenAI tools significantly enhance lower-order cognitive outcomes, particularly in understanding and applying concepts, with instructed use producing stronger positive effects than unguided use . However, their impact on higher-order cognitive skills such as creating and evaluating was minimal . This pattern suggests that AI effectively reduces intrinsic load for basic cognitive tasks but may not support—and potentially undermines—the productive mental effort required for deeper learning.

### Neural Evidence for Cognitive Engagement Changes

The study by Kosmyna et al. provides unique neurophysiological evidence for cognitive offloading effects. Using EEG measurements, they found that Brain-only participants exhibited the strongest, most distributed neural networks during essay writing, while LLM users displayed the weakest connectivity . Cognitive activity scaled down in relation to external tool use, with Search Engine users showing moderate engagement between these extremes . Critically, when LLM users were reassigned to write without AI assistance (LLM-to-Brain condition), they showed reduced alpha and beta connectivity indicating under-engagement . This neural evidence directly demonstrates that AI use can lead to measurable reductions in cognitive effort that persist even when tools are removed.

### The Problem of Cognitive Underload

Several studies identify a paradoxical risk: AI tools may reduce cognitive load below beneficial thresholds. Rohilla et al. frame this through Cognitive Load Theory, arguing that uncontrolled AI use can lead to cognitive underload, reducing effortful processing essential for deep learning . High AI users in their study showed 17.3 percentage points lower critical thinking scores and 22% fewer concepts retained compared to low users . The authors suggest this occurs because AI can minimize extraneous load but simultaneously eradicate desirable challenges necessary for skill development .

This concern is echoed by Fu et al., who observed that while AI tools initially facilitate higher-order thinking, students tend to converge toward passive reading engagement over time . Without deliberate design features, students default to superficial uses of AI, reducing desirable difficulties and effortful learning .

## Productivity Versus Learning: Evidence for Divergence

A central tension emerging across studies concerns the distinction between task completion efficiency and genuine learning outcomes. Multiple sources provide evidence that productivity gains from AI assistance may come at the cost of deeper learning.

### Documented Productivity-Learning Trade-offs

| Study | Productivity Effects | Learning Effects | Trade-off Evidence |
|---|---|---|---|
| Nataliya Kosmyna et al., 2025 | Immediate convenience, reduced friction | Underperformance at neural, linguistic, behavioral levels | Weaker memory traces, reduced self-monitoring |
| Dr. Atiya Rohilla et al., 2025 | Increased efficiency and task completion | Decreased critical thinking and memory retention | Paradox: tools for learning damage facilitated skills |
| Xiaodong Qu et al., 2025 | Effective for lower-order cognitive skills | Minimal impact on higher-order skills | Tension between quick completion and deeper engagement |
| Yaoying Han et al., 2025 | 28-35% improvement in technical accuracy | 25% decline in argument originality | Over-reliance creates ”cognitive crutches” |
| Viktoriia Pryma et al., 2025 | Reduced grammatical errors, enhanced vocabulary | May compromise cognitive effort and critical thinking | Surface-level improvement without deeper skill development |
| Yue Fu et al., 2025 | Quick comprehension and summarization | Passive reading engagement over time | Convenience undermines deeper learning |

Kosmyna et al.'s longitudinal findings are particularly striking: over four months, LLM users consistently underperformed at neural, linguistic, and behavioral levels compared to those using only their own cognitive resources . Despite the immediate convenience offered by AI tools, these users developed weaker memory traces, reduced self-monitoring capacities, and fragmented authorship experiences . Self-reported ownership of essays was lowest in the LLM group , and these users struggled to accurately quote their own work .

### The Quantitative Learning Deficit

Several studies quantify the learning costs associated with AI dependence. Rohilla et al. found highly relevant correlations between high AI dependency (≥5 hours/week) and measurable cognitive declines: critical thinking scores were 17.3 percentage points lower and memory retention was 22% worse among high AI users compared to low users . These differences were confirmed through t-tests and regression analysis .

Han et al.'s 18-week longitudinal study revealed that structured AI integration improved grammar and citation accuracy by 32% but correlated with a 19% decline in argument originality . Students in the unrestricted AI access group (Group A) showed a 25% decline in argument originality and 41% spent 35% less time refining arguments . These students also demonstrated a 15% decrease in their ability to articulate their writing processes, indicating atrophy of metacognitive skills .

### When Productivity and Learning Align

Not all studies found trade-offs between productivity and learning. The quasi-experimental study by Hong et al. demonstrated that AI-enabled cognitive offload instruction could enhance critical thinking while also improving essay quality in terms of logical coherence, evidence use, and originality . The key differentiating factor appears to be intentional instructional design: the intervention deliberately delegated lower-order writing tasks to AI while focusing student attention on analysis, evaluation, and reflection .

Similarly, Zhang et al. found that guided LLM use significantly enhanced both writing quality and academic engagement compared to unguided use or no use . Students in the guided condition achieved large effect sizes for writing quality and moderate effect sizes for engagement . In contrast, unguided use yielded only moderate gains in writing quality and did not produce significant effects on engagement or well-being .

## Effects on Skill Development and Retention

### Evidence for Skill Atrophy

Multiple studies document concerning patterns of skill atrophy associated with AI tool use. The evidence is particularly strong for higher-order cognitive skills including critical thinking, metacognition, and independent problem-solving.

| Study | Skills Affected | Direction of Effect | Magnitude | Mechanism |
|---|---|---|---|---|
| Nataliya Kosmyna et al., 2025 | Memory, self-monitoring, authorship | Decline | Neural connectivity differences | Reduced active integration of information |
| Dr. Atiya Rohilla et al., 2025 | Critical thinking, memory retention, problem-solving | Decline | 17.3% CT, 22% memory | Cognitive offloading, solution paralysis |
| Yaoying Han et al., 2025 | Argument originality, metacognitive skills | Decline | 19% originality, 15% metacognition | Over-reliance creating "cognitive crutches" |
| Hong Zhang et al., 2025 | Voice agency, originality | Potential decline | 82% voice appropriation concerns | Cognitive conflicts, textual homogenization |

| Study | Skills Affected | Direction of Effect | Magnitude | Mechanism |
|---|---|---|---|---|
| Viktoriia Pryma et al., 2025 | Independent writing, critical thinking | Surface improvement, depth decline | Not quantified | Passive acceptance of AI suggestions |

Rohilla et al. identified a particularly concerning phenomenon they term ”solution paralysis”—students' inability to engage in problem-solving without AI assistance . Their qualitative data revealed that students become less motivated for deep learning over time and experience increasing difficulty functioning without technological support . The authors interpret this through Digital Dependency theory, suggesting that unrestrained AI use can cause degradation of basic cognitive abilities due to decreased expenditure of mental effort .

### Disciplinary Differences in Skill Effects

The impact of AI on skill development varies across academic disciplines. Rohilla et al. found that humanities students recorded the sharpest cognitive drops from AI use , while Han et al. observed that STEM students benefited more from AI's technical support whereas humanities students reported loss of narrative voice authenticity . This suggests that disciplines requiring more interpretive, creative, or argumentative work may be particularly vulnerable to the negative cognitive effects of AI assistance.

### Evidence for Skill Development

Not all findings indicate skill decline. Several studies found that with appropriate pedagogical design, AI tools can support skill development. Hong et al. found significant improvements in standardized critical thinking assessments among students in the AI-enabled cognitive offload group . The intervention enhanced analytical thinking, problem-solving, and effective communication while creating an educational environment supporting originality and responsible writing practices .

Xiao et al. demonstrated that training students in ”pedagogical prompting”—constructing prompts that elicit learning-oriented rather than answer-providing responses—led to significant learning gains in students' ability to use AI effectively for learning rather than task completion . Effect sizes ranged from 0.82 to 0.86 across six components of pedagogical prompts ($p < .001$) , and students showed increased willingness to use these approaches in future AI interactions .

### Transfer and Independence Concerns

A critical gap in the literature concerns transfer of learning to non-AI contexts. Most studies do not assess whether skills developed with AI assistance transfer to unassisted contexts . Kosmyna et al.'s crossover design provides indirect evidence that such transfer may be limited: when LLM users were reassigned to write without assistance, they showed reduced neural connectivity indicative of under-engagement rather than successful independent performance .

The development of AI-independent capabilities remains underexplored. Pryma et al. suggest that AI should be used as a supplement rather than substitute to maintain independent writing abilities , while Han et al. emphasize the importance of maintaining human agency and critical thinking even when using AI tools . The systematic review by Zaheer et al. explicitly identifies the need for balanced pedagogical practices that foster independent writing skills .

## Engagement Patterns and Usage Behaviors

### Evolution of AI Usage Over Time

Studies tracking student engagement over time reveal concerning patterns of increasing passivity and dependency. Fu et al. found that while students initially engaged in higher-order thinking (Analyzing and Evaluating prompts), they converged toward passive reading engagement over subsequent weeks, relying more heavily on understanding-level prompts . This pattern was consistent: students defaulted to surface-level use for quick comprehension rather than strategic use for deeper inquiry .

| Study | Initial Pattern | Pattern Over Time | Trajectory |
|---|---|---|---|
| Yue Fu et al., 2025 | Higher-order thinking (Analyzing, Evaluating) | Passive reading engagement | Declining cognitive engagement |
| Hong Zhang et al., 2025 | High dependency (4.33 uses/session) | Critical avoidance (1.49 uses) | From dependency to avoidance |
| Yaoying Han et al., 2025 | Heavy reliance on AI for structure | Decline in metacognitive skills | Declining self-regulation |
| Dr. Atiya Rohilla et al., 2025 | Use for immediate solutions | Increased dependency, solution paralysis | Deepening dependency |

Hong Zhang et al. documented an interesting trajectory: early technology dependency (averaging 4.33 self-initiated uses per session) shifted to critical avoidance (1.49 uses) as learners confronted AI's limitations . However, this avoidance emerged after students had already experienced cognitive conflicts in reconciling AI outputs with their original ideas, with 82% reporting voice appropriation concerns .

### Active Versus Passive Engagement

The distinction between active and passive engagement with AI tools emerges as a critical factor across studies. Kosmyna et al. found clear neural signatures differentiating engagement modes: Brain-only participants showed broad, distributed neural networks indicative of active cognitive processing, while LLM users displayed weak connectivity patterns consistent with passive information consumption .

Arif et al. observed that students engaged in revision cycles with AI feedback but also displayed superficial acceptance of AI suggestions . Some students showed reduced willingness to engage in independent editing over time, relying heavily on AI suggestions rather than developing their own editorial judgment .

The programming education literature reveals particularly concerning patterns. Arora et al. found overwhelming adoption of LLMs with many students copying entire assignment descriptions for complete solutions . While students used both strategic (iterative refinement) and surface-level (copying assignment descriptions) approaches, the trend toward over-reliance was prevalent .

### Self-Regulation and Metacognitive Behaviors

Critical thinking disposition appears to be a key moderator of how students engage with AI tools. Hou et al. found that critical thinking is positively associated with reflective, collaborative, and cautious use of generative AI, while high trust without critical thinking leads to thoughtless use . Importantly, critical thinking can offset the influence of trust on dependency behaviors and amplify the positive effects of AI literacy on beneficial usage patterns .

Pryma et al. found that students showed passive engagement by accepting AI suggestions without fully understanding them . Students lacked self-regulation by not thoroughly checking AI corrections, instead relying on the AI's apparent authority . This pattern raises concerns about the development of critical evaluation skills essential for professional competence.

### Over-reliance Indicators

Multiple studies document indicators of over-reliance and dependency:

- Students copying entire assignment descriptions for complete solutions
- Inability to accurately quote one's own AI-assisted work
- Solution paralysis when AI tools are unavailable
- Reduced time spent on pre-writing activities and argument refinement
- Early technology dependency averaging 4+ uses per session
- Preference for AI assistance over human support for academic tasks

Crawford et al. found that students increasingly substituted AI for human social support in academic contexts, which was associated with negative effects on achievement when psychological wellbeing and sense of belonging were considered . This suggests that AI dependency may have broader effects on student academic experience beyond cognitive outcomes.

## The Moderating Role of Instructional Design

### Guided Versus Unguided Use

The most consistent finding across studies concerns the critical importance of pedagogical scaffolding in determining whether AI use supports or undermines learning. The contrast between guided and unguided use appears across multiple domains and measures.

| Study | Guided Use Effects | Unguided Use Effects | Difference |
|---|---|---|---|
| Xiaodong Qu et al., 2025 | Stronger positive effects on lower-order skills | Weaker effects | Instructional context significantly moderates effect sizes |
| Min Zhang et al., 2025 | Large effect on writing quality, moderate on engagement | Moderate gains in writing, no effect on engagement | Structured guidance essential for full benefits |
| Hareem Arif et al., 2025 | Enhanced CT with scaffolding | Suppressed independent analysis without scaffolding | Task design decisive for outcomes |
| Kazemitabaar et al., 2024 | CodeAid: 87% correctness, 91% helpfulness | ChatGPT concerns about direct solutions | Guardrails support learning engagement |

The meta-analysis by Qu et al. found that instructional context significantly moderates effect sizes, with guided use producing stronger positive effects than unguided use . Zhang et al.'s experimental comparison directly demonstrated that guided LLM use significantly improved academic writing quality and academic engagement compared to unguided use, with unguided use showing limited impact on mental health and engagement despite some writing improvements .

### Effective Scaffolding Strategies

Studies identify several strategies that appear to support beneficial AI use:

**Cognitive offload instruction** : Hong et al. found that deliberately delegating lower-order tasks to AI while focusing student attention on higher-order thinking led to significant critical thinking improvements . The intervention involved structured writing cycles including AI brainstorming, individual critique, peer-AI co-revision, and reflective journaling .

**Phased implementation** : Han et al. recommend limiting AI use to post-draft stages to maintain pre-writing cognitive engagement . Their ”co-creative framework” emphasizes transparency, phased tool usage, and adaptive assessment to preserve human agency .

**Pedagogical prompting** : Xiao et al. developed training in constructing prompts that elicit learning-oriented rather than answer-providing responses from AI . This approach led to significant gains in students' ability to use AI for genuine learning support .

**Strategic scaffolding with metacognitive wrappers** : Rohilla et al. recommend ”strategic scaffolding, metacognitive wrappers, and cognitive load calibration” as evidence-based approaches for sustainable AI implementation . These approaches aim to maintain cognitive engagement while leveraging technological opportunities.

**Guardrails against direct solutions** : The CodeAid system was specifically designed to provide helpful, technically correct responses without revealing direct code solutions . This approach used pseudo-code with line-by-line explanations and annotated student code with fix suggestions rather than providing complete solutions . Educators interviewed in the study emphasized the importance of these guardrails for maintaining learning engagement .

### Institutional and Instructor Roles

Several studies highlight the importance of instructor involvement in mediating AI tool effects. Maphalala et al.'s systematic review found that academics' perspectives on AI adoption vary based on technological proficiency, pedagogical beliefs, and institutional support . Successful AI integration requires alignment with pedagogical theories such as constructivism, connectivism, and self-directed learning .

Zulaikha et al. found that students valued teacher feedback alongside AI tools and developed ethical awareness regarding authorship and originality through guided instruction . Despite AI assistance, human revision remained necessary for semantic precision and rhetorical fit, especially in specialized content .

## Synthesis: Reconciling Conflicting Findings

The evidence across these 25 studies presents an apparent contradiction: some studies find AI tools enhance learning outcomes while others document significant cognitive costs. This heterogeneity can be explained through several systematic factors.

## Context and Implementation Distinctions

The divergent findings map systematically onto implementation characteristics. Studies finding positive cognitive effects consistently employed structured scaffolding, deliberate pedagogical design, and guided use protocols . Studies finding negative effects typically examined free use or unstructured access conditions .

Specifically:

- **Positive outcomes** occurred with: instructor-guided use , structured prompts and workflows , explicit training in AI interaction , and phased implementation limiting AI to specific task stages
- **Negative outcomes** occurred with: free use without guidance , extended duration of unstructured access , and high-intensity use (≥5 hours/week)

## Cognitive Skill Level as Moderator

The meta-analysis by Qu et al. provides a critical moderating variable: cognitive skill level targeted by the task. GenAI tools significantly enhance lower-order cognitive outcomes (understanding, applying) with an overall effect size of 1.34 (95% CI: 1.17, 1.50) . However, impact on higher-order skills (analyzing, evaluating, creating) was minimal . This suggests both findings may be correct within their respective margins:

- AI assistance **benefits** learning when tasks require comprehension, application, and procedural knowledge
- AI assistance **fails to benefit or harms** learning when tasks require analysis, synthesis, evaluation, and creative thinking

## Temporal Dynamics

Studies with different measurement timeframes yield different conclusions. Immediate assessments tend to show productivity benefits and performance improvements . Longitudinal studies reveal concerning patterns of skill atrophy, declining engagement, and dependency development .

The Kosmyna et al. study is particularly instructive: over four months of observation, the cognitive costs of AI use accumulated rather than diminished . This suggests that cross-sectional or short-term studies may systematically underestimate the cognitive risks of AI tool use while capturing immediate productivity benefits.

## Methodological Quality Hierarchy

Weighting evidence by methodological rigor reveals that studies with stronger designs tend toward more cautious conclusions. The only study employing neuroimaging found clear evidence of reduced cognitive engagement in AI users . The meta-analysis, aggregating across multiple studies with quantitative outcomes, found null effects for higher-order skills despite positive effects for lower-order skills .

Studies relying primarily on self-report measures or satisfaction ratings tend to show more positive outcomes , while studies with behavioral measures, cognitive assessments, or neural measures reveal more concerning patterns .

## Mechanistic Explanation

The divergent findings can be integrated through cognitive load theory as an explanatory mechanism:

1. **AI reduces extraneous load** by streamlining information access and handling mechanical tasks —this is consistently beneficial

2. **AI reduces intrinsic load** by simplifying task demands —this can be beneficial for novices or genuinely excessive demands
3. **AI can eliminate germane load** when it removes the productive struggle necessary for skill development —this produces the negative learning outcomes observed in unstructured use conditions

The key insight is that cognitive effort is not uniformly harmful to learning. ”Desirable difficulties” that increase immediate task demands often enhance long-term retention and skill development. When AI eliminates these beneficial challenges, short-term performance improves while long-term learning suffers. This explains why productivity metrics and learning outcomes can diverge, and why immediate satisfaction does not predict educational value.

### Population and Context Specificity

The effects of AI on learning are not uniform across populations or contexts:

- **Disciplinary differences** : Humanities students show greater cognitive declines than STEM students , while STEM students benefit more from AI's technical support
- **Prior skill level** : Students with stronger foundational skills are more likely to use AI effectively , while less skilled students may avoid AI or use it unproductively
- **Critical thinking disposition** : Students high in critical thinking show reflective, cautious, and collaborative AI use, while those with high trust but low critical thinking show thoughtless use

### Synthesis Conclusions

The evidence supports several specific conclusions rather than a simple positive or negative verdict:

1. **AI tools reliably improve immediate task performance and productivity** across contexts, but this improvement does not automatically translate to learning gains
2. **AI tools effectively support lower-order cognitive skills** (comprehension, application) but provide minimal benefit for higher-order skills (analysis, evaluation, creation)
3. **Unstructured, high-intensity AI use leads to measurable cognitive costs** including reduced critical thinking, memory retention, and metacognitive skills
4. **Structured, guided AI use can support learning** when design deliberately maintains cognitive engagement and reserves higher-order tasks for human effort
5. **Temporal patterns matter** : benefits appear immediately while costs accumulate over extended use
6. **Individual differences in critical thinking disposition** strongly moderate whether AI use supports or undermines learning

## Methodological Limitations and Research Gaps

### Sample and Design Limitations

Most studies have limited generalizability due to sample constraints. Many studies were conducted in single institutions , specific geographic contexts , or with homogeneous populations . The meta-analysis by Qu et al. notes that its sample of only 17 studies may lead to bias and limited generalizability .

Duration limitations are pervasive. While Kosmyna et al. tracked participants over four months and Han et al. conducted an 18-week study , most studies capture only immediate or short-term effects . Rohilla et al. explicitly note that their six-month observation period was insufficient to observe long-term effects , and multiple authors call for longitudinal research to assess sustained impacts .

### Measurement Challenges

A significant gap exists in cognitive measurement approaches. Only one study employed neuroimaging , providing direct evidence of cognitive engagement changes. Most studies rely on self-report measures susceptible to recall bias and social desirability effects . The reliance on self-reported data is explicitly acknowledged as a limitation by multiple authors .

Assessment of AI's effects on learning (as opposed to task performance) remains challenging. Studies often use productivity metrics or satisfaction ratings rather than validated learning outcome measures . Pryma et al. note limited empirical analysis of AI's influence on specific linguistic features across different proficiency levels , while Zaheer et al. identify limited effects on advanced critical thinking as indicating a gap in cognitive measures .

### Critical Research Gaps

Several important questions remain inadequately addressed:

**Transfer of learning** : Whether skills developed with AI assistance transfer to unassisted contexts is largely unexplored . The evidence suggesting reduced neural engagement when AI users attempt unassisted tasks raises concerns that such transfer may be limited.

**Long-term cognitive effects** : Extended longitudinal studies tracking cognitive development over years rather than weeks or months are absent from the literature. Authors consistently call for such research .

**Optimal scaffolding design** : While multiple studies identify scaffolding as critical, specific evidence on optimal scaffolding designs for different contexts, tasks, and populations is lacking .

**Neurological mechanisms** : Beyond Kosmyna et al.'s study, no research has examined the neural mechanisms underlying AI's cognitive effects. This represents a major gap given the clear theoretical relevance of neurocognitive evidence to questions of learning and skill development.

**Equity implications** : While Han et al. note disparities in AI tool access between socioeconomic groups , systematic investigation of how cognitive offloading effects vary by student resources, prior preparation, or educational context is minimal.

### Gaps Between Productivity and Learning Assessment

A fundamental methodological issue concerns the conflation of productivity metrics with learning outcomes. Studies that assess AI effects using immediate task performance, satisfaction ratings, or writing quality scores may systematically miss learning costs that manifest only over time or in transfer contexts.

Bernik et al. explicitly acknowledge that their research on AI-assisted grading does not address cognitive offloading effects or the impact on learning outcomes and skill development . This gap between what is commonly measured (productivity, performance, satisfaction) and what matters for education (learning, retention, transfer, skill development) pervades the literature and limits confidence in conclusions about AI's educational value.

The field would benefit from studies that simultaneously assess immediate performance and delayed retention, assisted performance and unassisted transfer, and satisfaction alongside validated cognitive skill measures. Such designs would enable more rigorous conclusions about when productivity gains align with or diverge from learning benefits.

## References


A. Bernik, D. Radošević, and Andrej Čep. "A Comparative Study of Large Language Models in Programming Education: Accuracy, Efficiency, and Feedback in Student Assignment Grading." *Applied Sciences*, 2025.

Alexandru Dinu. "The Influence of AI-Assisted Tools on Engineering Project Outcomes." *Revista Romaneasca Pentru Educatie Multidimensionala*, 2025.

Chenyu Hou, Gaoxia Zhu, and Vidya Sudarshan. "The Role of Critical Thinking on Undergraduates' Reliance Behaviours on Generative AI in Problem-solving." *British Journal of Educational Technology*, 2025.

Dr.Atiya Rohilla. "Impact of Excessive AI Tool Usage on the Cognitive Abilities of Undergraduate Students: A Mixed Method Study." *Advance Social Science Archive Journal*, 2025.

Hardiyanti Pratiwi, Suherman, Hasruddin, and Muhammad Ridha. "Between Shortcut and Ethics: Navigating the Use of Artificial Intelligence in Academic Writing Among Indonesian Doctoral Students." *European Journal of Education*, 2025.

Hareem Arif, and Javairia Naeem. "THE IMPACT OF GENERATIVE AI ON LEARNER AUTONOMY AND CRITICAL THINKING IN ENGLISH AS A FOREIGN LANGUAGE (EFL) WRITING CLASSROOMS." *Journal of Applied Linguistics and TESOL (JALT)*, 2025.

Hong Zhang. "An Observational Study on Challenges of Using AI Tools in EFL Writing." *English Language Teaching*, 2025.

Hui Hong, C. Viriyavejakul, and P. Vate-U-Lan. "Enhancing Critical Thinking Skills: Exploring Generative AI-Enabled Cognitive Offload Instruction in English Essay Writing." *Journal of Ecohumanism*, 2025.

Hui Hong, P. Vate-U-Lan, and C. Viriyavejakul. "Cognitive Offload Instruction with Generative AI: A Quasi-Experimental Study on Critical Thinking Gains in English Writing." *Forum for Linguistic Studies*, 2025.

Joseph Crawford, Kelly-Ann Allen, Bianca Pani, and Michael Cowling. "When Artificial Intelligence Substitutes Humans in Higher Education: The Cost of Loneliness, Student Success, and Retention." *Studies in Higher Education*, 2024.

M. Maphalala, and O. A. Ajani. "Leveraging Artificial Intelligence as a Learning Tool in Higher Education." *Interdisciplinary Journal of Education Research*, 2025.

M. Mutanga, Jotham Msane, Thaddeus Ndumiso Mndaweni, Bongokuhle Brightman Hlongwane, and Neliswa Ziyanda Ngcobo. "Exploring the Impact of LLM Prompting on Students' Learning." *Trends in Higher Education*, 2025.

Majeed Kazemitabaar, Runlong Ye, Xiaoning Wang, Austin Z Henley, Paul Denny, Michelle Craig, and Tovi Grossman. "CodeAid: Evaluating a Classroom Deployment of an LLM-Based Programming Assistant That Balances Student and Educator Needs." *International Conference on Human Factors in Computing Systems*, 2024.

Min Zhang. "Optimizing Academic Engagement and Mental Health Through AI: An Experimental Study on LLM Integration in Higher Education." *Frontiers in Psychology*, 2025.

Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. "Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task." *arXiv.org*, 2025.

Ruiwei Xiao, Xinying Hou, Runlong Ye, Majeed Kazemitabaar, Nicholas Diana, Michael Liut, and John Stamper. "Improving Student-AI Interaction Through Pedagogical Prompting: An Example in Computer Science Education." *arXiv.org*, 2025.

Samuel Boguslawski, Rowan Deer, and Mark G. Dawson. "Programming Education and Learner Motivation in the

Age of Generative AI: Student and Educator Perspectives." *Information and Learning Sciences*, 2024.
Sanka Rasnayaka, Guanlin Wang, Ridwan Shariffdeen, and Ganesh Neelakanta Iyer. "An Empirical Study on Usage and Perceptions of LLMs in a Software Engineering Project." *2024 IEEE/ACM International Workshop on Large Language Models for Code (LLM4Code)*, 2024.
Sidra Zaheer, Congzhi Ma, Yimeng Zhu, and Sheri Vasinda. "GenAI in Academic Writing- Empowering Learners or Redefining Traditional Pedagogical Practices?" *International Journal of Artificial Intelligence*, 2025.
Utkarsh Arora, Anupam Garg, Aryan Gupta, Samyak Jain, Ronit Mehta, Rupin Oberoi, Prachi, et al. "Analyzing LLM Usage in an Advanced Computing Class in India." *Proceedings of the 27th Australasian Computing Education Conference*, 2024.
Viktoriia Pryma, Oksana Pelivan, Tetiana Teletska, Olha Tsobenko, and Natalia Zagrebelna. "AI Writing Assistants and Student Competence: A Linguistic Aspect." *Arab World English Journal*, 2025.
Xiaodong Qu, Joshua Sherwood, Peiyan Liu, and Nawwaf Aleisa. "Generative AI Tools in Higher Education: A Meta-Analysis of Cognitive Impact." *CHI Extended Abstracts*, 2025.
Yaoying Han. "Beyond the Algorithm: Reconciling Generative AI and Human Agency in Academic Writing Education." *International Journal of Learning and Teaching*, 2025.
Yue Fu, and Alexis Hiniker. "Supporting Students' Reading and Cognition with AI." *arXiv.org*, 2025.
Zulaikha Zulaikha, Cahya Edi Setyawan, M. Mabruri, and Siti Rauhillah. "The Effectiveness of Artificial Intelligence and Deep Learning Tools in Enhancing Academic Journal Writing: A Mixed Methods Study of Arabic Language Education Students in Indonesia." *Transformasi*, 2025.