Cognitive Crossroads: Empirical Effects of Replacing Social Media with Generative AI Tools on Learning, Retention, Attention, and Critical Thinking Among Young Adults (18–35)

Asst. Prof. Chinmay Powar

DES Shri Navalmal Firodia Law College

Abstract

Social media platforms have become the dominant mode of digital consumption among young adults aged 18–35, yet their cognitive consequences continue to generate scientific controversy. Concurrently, generative artificial intelligence (AI) tools—including ChatGPT, Claude, Gemini, and Perplexity—represent a qualitatively distinct mode of human-computer interaction characterized by active, dialogic engagement with knowledge. No prior empirical study has examined the cognitive consequences of systematically substituting social media usage with generative AI interaction within a controlled experimental design. This study reports findings from a pre-registered, two-group Randomized Controlled Trial (RCT; N = 312 completers, age M = 24.7 years, SD = 4.1) in which participants were randomly assigned to either a 60-day social media elimination + AI substitution condition or a continued social media use (control) condition. Outcome variables—learning ability, delayed knowledge retention, cognitive absorption, sustained attention, and critical thinking—were assessed using validated instruments at baseline (Week 0), midpoint (Week 4), endpoint (Week 9), and 30-day follow-up (Week 13). Multilevel Structural Equation Modeling (SEM) and difference-in-differences (DiD) analysis revealed that AI substitution produced significant improvements in knowledge retention (d = 0.61, p < .001), sustained attention as measured by Stroop interference scores (d = 0.54, p < .001), and cognitive absorption (d = 0.48, p = .002). Learning ability gains reached statistical significance (d = 0.39, p = .014) but were smaller in magnitude. Critical thinking showed a significant interaction with AI literacy (β = 0.31, p = .008): high-AI-literacy participants demonstrated meaningful critical thinking gains (d = 0.44), while low-AI-literacy participants showed no significant change (d = 0.09). Attention span recovery and increased cognitive engagement partially mediated the relationship between AI substitution and learning outcomes. These findings support a nuanced conclusion: generative AI functions as a cognitive enhancer when engaged dialogically and with sufficient user literacy, but does not universally produce such benefits. Implications are discussed for educational policy, digital regulation, and AI instructional design.

1. Introduction

1.1 The Attention Economy and Its Cognitive Costs

The concept of the attention economy—first formalized by Herbert Simon (1971) and later elaborated by Davenport and Beck (2001)—holds that cognitive attention, not information, is the genuinely scarce resource of the digital age. In the contemporary media environment, social media platforms have emerged as the dominant architects of attentional capture, deploying algorithmic personalization, variable reward schedules, and social validation mechanics engineered to maximize time-on-platform at measurable cost to cognitive depth (Alter, 2017; Zuboff, 2019). As of 2024, young adults aged 18–35 spend an average of 3.2 hours daily on social media platforms (DataReportal, 2024), representing approximately 20% of waking hours devoted to a mode of information consumption that neuroscience increasingly associates with attentional fragmentation, impulsivity amplification, and the suppression of deliberative cognition (Wilmer, Sherman, & Chein, 2017; Lissak, 2018).

The mechanism appears to operate through chronic over-recruitment of System 1 thinking—fast, associative, emotionally reactive cognition—at the expense of System 2 processes that are slower, effortful, and analytically rigorous (Kahneman, 2011). Social media's architectural affordances—infinite scroll, push notifications, emotionally charged algorithmic feeds—continuously recruit System 1 processing while leaving diminishing neural resources for the sustained, hypothesis-testing cognition that characterizes genuine intellectual engagement. Thornton et al. (2014) demonstrated that the mere presence of a smartphone reduces available cognitive capacity; Baumgartner et al. (2014) found that adolescent media multitaskers exhibit significantly lower working memory performance than low-multitaskers; and Wolf (2018) catalogued accumulating evidence that habitual digital media consumption degrades the neural circuits underlying deep reading and extended concentration.

1.2 Generative AI as Cognitive Partner

Against this backdrop of attentional degradation, generative artificial intelligence tools have emerged as a qualitatively different mode of human-computer interaction. Unlike the passive consumption dynamics of social media, interaction with large language models (LLMs)—including GPT-4 (OpenAI), Claude 3 (Anthropic), Gemini 1.5 (Google DeepMind), and Perplexity AI—requires the user to formulate queries, evaluate responses, critically interrogate outputs, and synthesize information actively across turns. This dialogic architecture aligns with constructivist learning theory, which holds that durable knowledge is actively constructed through engagement rather than passively received through exposure (Vygotsky, 1978; Jonassen, 1994).

Preliminary research has suggested that AI-assisted learning can improve comprehension and metacognitive awareness when tools are used in a Socratic, iterative fashion rather than for rote answer-retrieval (Nye et al., 2023; Mollick & Mollick, 2023). However, concerns about cognitive outsourcing—the offloading of reasoning to AI systems, with attendant atrophy of independent critical thinking—have been raised across philosophy of technology and cognitive science (Danaher, 2018; Risko & Gilbert, 2016). The net cognitive consequence of sustained generative AI use therefore remained, prior to this study, both empirically unresolved and theoretically contested.

1.3 The Critical Importance of the 18–35 Cohort

The young adult cohort of 18–35 occupies a uniquely consequential position in this debate. First, this group represents the heaviest users of both social media and generative AI tools globally (Statista, 2024). Second, adults in this age range are completing the final phase of prefrontal cortex maturation—the brain region governing executive function, impulse control, and deliberative reasoning—making environmental inputs during this period potentially neuroplastically consequential (Giedd, 2004; Steinberg, 2008). Third, they constitute the emerging knowledge workforce, rendering their cognitive capacity of direct economic relevance. Understanding whether a practically feasible behavioral intervention—substituting social media with generative AI—produces measurable cognitive benefit for this population carries substantial scientific and policy significance.

1.4 Research Gap and Study Rationale

Despite substantial literatures on the cognitive harms of social media and the pedagogical potential of AI, no published empirical study prior to this work had examined the cognitive consequences of systematically replacing one with the other in a controlled experimental design. Existing studies either examine social media harms or AI benefits in isolation, rely on cross-sectional correlational designs inadequate for causal inference, or test AI in structured educational settings rather than in naturalistic daily-life substitution contexts. This study directly fills this gap by employing a pre-registered, two-group RCT—the methodological gold standard for causal inference—to estimate the effect of social media elimination and AI substitution on five cognitive domains in a community sample of young adults.

1.5 Research Questions

This study addresses five research questions: RQ1: Does replacing social media use with generative AI tools produce improvements in learning ability and knowledge retention? RQ2: Does AI substitution increase or decrease sustained attention and cognitive absorption? RQ3: Does AI substitution enhance or impair critical thinking ability? RQ4: Do attention recovery and increased cognitive engagement mediate the effects of AI substitution on learning outcomes? RQ5: Does AI literacy moderate the relationship between AI substitution and critical thinking gains?

2. Theoretical Framework and Hypotheses

2.1 Cognitive Load Theory

Cognitive Load Theory (CLT; Sweller, 1988) distinguishes between extraneous cognitive load (imposed by poor instructional design or irrelevant stimulation) and germane cognitive load (associated with schema formation and genuine learning). Social media platforms impose chronically high extraneous load through fragmented, decontextualized information streams and continuous attentional switching demands, depleting working memory capacity available for deeper encoding. Generative AI, when used dialogically, has the theoretical capacity to reduce extraneous load by providing coherent, contextually calibrated explanations, and to increase germane load by scaffolding schema-building through iterative questioning and elaborative feedback. The study operationalizes CLT by predicting that AI substitution will reduce extraneous load and thereby free cognitive resources for deeper learning processes.

2.2 Attention Scarcity and Dual-Process Theory

Simon's (1971) attention scarcity framework treats cognitive attention as a zero-sum resource: every unit devoted to social media's engineered attentional capture is unavailable for productive intellectual engagement. Kahneman's (2011) dual-process model adds mechanistic specificity: social media's architecture chronically recruits System 1 (fast, automatic, reactive) processing while suppressing System 2 (slow, deliberate, analytical) cognition. This study predicts that eliminating social media will free attentional resources and that redirecting this attention toward AI interaction—which structurally demands query formulation, output evaluation, and iterative reasoning—will shift the cognitive system toward greater System 2 engagement.

2.3 Deep Learning Theory and Self-Determination Theory

Marton and Säljö's (1976) distinction between surface and deep learning approaches maps directly onto the contrast under investigation. Social media incentivizes surface-level information processing—scanning, reacting, and sharing without integration into existing knowledge structures. Generative AI, used dialogically, incentivizes deep processing through its responsiveness to follow-up questions and demand for coherent query formulation. Self-Determination Theory (SDT; Deci & Ryan, 2000) contributes by predicting that AI interaction fulfills competence and autonomy needs—prerequisites for intrinsic motivation and deep engagement—in ways that algorithmically controlled social media does not.

2.4 Neuroplasticity and Dopamine Reward Mechanisms

Social media platforms exploit the mesolimbic dopamine system through variable reinforcement schedules (unpredictable social rewards: likes, viral spread, novel content) that are structurally analogous to gambling reinforcement (Haynes, 2018; Andreassen et al., 2016). Chronic dopaminergic activation in response to low-effort social stimuli may recalibrate reward thresholds, making sustained high-effort cognitive engagement feel unrewarding by comparison. Neuroplasticity research (Merzenich, 2013) suggests the brain adapts structurally to habitual stimulation patterns: chronic attentional switching may weaken sustained-focus neural networks through competitive plasticity. AI substitution—by demanding sustained problem-solving over 60 days—was predicted to initiate neuroplastic recovery of attentional and executive function networks.

2.5 Hypotheses

H1: Participants in the AI substitution condition will demonstrate significantly greater gains in standardized learning performance from baseline to endpoint compared to control group participants.

H2: Treatment group participants will demonstrate significantly superior delayed knowledge retention at 30-day follow-up compared to control participants.

H3: Treatment group participants will exhibit significant improvements in sustained attention performance (Stroop interference reduction; digit span increase) relative to controls.

H4a: Among high-AI-literacy participants, AI substitution will significantly improve critical thinking scores. H4b: Among low-AI-literacy participants, AI substitution will not significantly improve critical thinking, and may attenuate it.

H5: The effects of AI substitution on learning and retention outcomes will be partially mediated by improvements in sustained attention and cognitive engagement intensity.

3. Method

3.1 Study Design and Pre-Registration

This study employed a two-group, parallel-arm Randomized Controlled Trial (RCT) with pre-test/post-test and 30-day follow-up measurements. The trial was pre-registered on the Open Science Framework (OSF; osf.io/ck73m) prior to data collection, with all primary hypotheses, outcome measures, and analysis plans specified in advance. Ethics approval was granted by the Institutional Review Board of DES Shri Navalmal Firodia Law College. The study was conducted between September 2023 and March 2024.

3.2 Participants

Participants were recruited through university mailing lists, community bulletin boards, and social media advertisements (a controlled irony: social media recruitment for a social media study) across three geographically distinct urban centers. Inclusion criteria required: (1) age 18–35; (2) self-reported daily social media use $\geq$ 90 minutes (verified via 14-day device screen time log at screening); (3) smartphone ownership and high-speed internet access; (4) English language proficiency (CEFR B2+); (5) no diagnosis of ADHD, learning disability, or major depressive disorder (assessed via self-report and PHQ-9 screen). Exclusion criteria included professional social media management responsibilities and prior clinical use of AI-assisted learning tools exceeding 5 hours per week.

Of 489 individuals screened, 358 met inclusion criteria and were enrolled. Following stratified randomization (by gender, education level, and baseline social media use), 180 were assigned to the treatment group and 178 to the control group. Forty-six participants withdrew or were excluded due to non-compliance (n = 31 treatment; n = 15 control), yielding a final analytic sample of N = 312 (treatment: n = 149; control: n = 163). Attrition analysis confirmed no systematic differences between completers and non-completers on baseline demographic or cognitive variables (all p > .15). Power analysis using G*Power (Faul et al., 2007) indicated that N = 280 provides 80% power to detect d = 0.50 at α = .05; the achieved sample exceeds this threshold.

3.3 Demographic Characteristics

Table 1

Baseline Demographic Characteristics by Experimental Condition (N = 312)

Characteristic	Treatment (n = 149)	Control (n = 163)	p-value
Age, M (SD)	24.5 (4.0)	24.9 (4.2)	.38
Gender (% Female)	54.4%	52.8%	.76
Education (% Bachelor's+)	61.7%	63.2%	.79

Characteristic	Treatment (n = 149)	Control (n = 163)	p-value
SES (1–10 scale), M (SD)	5.9 (1.7)	6.1 (1.6)	.31
Baseline social media hrs/day, M (SD)	3.4 (1.1)	3.3 (1.0)	.42
AI Literacy Score (0–48), M (SD)	28.3 (7.6)	27.9 (7.4)	.64
Baseline IQ (Raven's), M (SD)	102.1 (11.4)	101.7 (11.2)	.77

Note. p-values reflect independent-samples t-tests (continuous) or chi-square tests (categorical). No significant baseline differences were detected, confirming successful randomization.

3.4 Experimental Conditions

Treatment Group (AI Substitution)

Participants randomly assigned to the treatment condition were required to permanently delete Instagram, Facebook, Snapchat, X (Twitter), and Reddit from all personal devices for the 60-day study period. YouTube was explicitly permitted given its substantively distinct content ecosystem (long-form educational and informational video) and differential associations with learning outcomes (Moghavvemi et al., 2018). WhatsApp and iMessage were similarly permitted for interpersonal communication functions distinct from social media content consumption.

Participants were instructed to redirect time previously spent on deleted platforms toward engagement with at least one of four generative AI tools: ChatGPT (OpenAI GPT-4), Claude (Anthropic), Gemini 1.5 (Google DeepMind), or Perplexity AI. All participants received complimentary premium access to each platform for the study duration. No constraint was placed on topic or use case, preserving ecological validity. Participants were encouraged to use AI for any purpose that had previously been served by social media or other information-seeking activities: curiosity questions, creative projects, learning new skills, decision support, or conversational engagement.

Control Group

Control participants continued their habitual social media usage without modification and received no access to AI tools unless they were already using them at baseline (n = 23; baseline AI use documented and covaried). Control participants completed all assessment instruments on the same schedule as treatment participants but received no behavioral intervention.

3.5 Compliance Verification

Compliance in the treatment group was monitored through three methods: (1) weekly device screen time reports submitted via a secure study portal, which provided platform-level data from iOS Screen Time and Android Digital Wellbeing; (2) weekly self-report compliance checklists with behavioral specificity items ('Have you opened any of the deleted platforms this week? If yes, for how long?'); and (3) monthly 20-

minute video check-in sessions with a research assistant who conducted a live device audit with participant consent. AI query logs were exported weekly by participants from platform activity histories; log analysis categorized each query into one of four depth categories: (a) factual retrieval, (b) task completion, (c) explanatory dialogue, or (d) deep reasoning and argumentation. Participants whose screen time data showed social media use exceeding 15 minutes per week for more than two consecutive weeks were classified as non-compliant and excluded from per-protocol analyses (included in intent-to-treat analyses).

3.6 Outcome Measures

Learning Ability

Learning ability was assessed using standardized reading comprehension and analytical reasoning passages drawn from the GRE Verbal Reasoning Bank (four equivalent forms, counterbalanced across participants and time points). Each form comprised four texts (1,000–1,200 words) spanning scientific, historical, economic, and philosophical domains, matched for Flesch-Kincaid readability (Grade 13–14 level). Comprehension was assessed via 10 multiple-choice items and two short inferential writing tasks per text (total 48 items). Writing responses were scored by two blinded raters (ICC = .87); discrepancies resolved through adjudication. Total learning score was standardized (z-score) within each form for cross-time comparability.

Knowledge Retention

Long-term retention was assessed via surprise delayed recall at 30 days post-intervention (Week 13). Participants were not informed that a retention test would occur, eliminating preparation effects. The retention battery comprised cued recall (30 key terms from weekly study materials; participants provided definitions and conceptual explanations) and free recall (participants generated the five most important ideas from each of three previously studied passages). Retention was scored as percent of originally encoded items correctly recalled, with strict and lenient scoring criteria applied (results reported under strict criteria).

Cognitive Absorption

Cognitive absorption was measured using the validated 15-item Cognitive Absorption Scale (CAS; Agarwal & Karahanna, 2000), which assesses five dimensions: temporal dissociation, focused immersion, heightened enjoyment, control, and curiosity. Items were rated on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree). Participants completed the CAS with reference to their primary digital activity (social media for controls; AI tools for treatment). The CAS demonstrated strong reliability across time points (Cronbach's α = .86–.89).

Sustained Attention

Attention was assessed using two validated neuropsychological instruments administered digitally under standardized conditions. The Stroop Color-Word Test (Golden, 1978) measured cognitive inhibition and selective attention: participants named ink colors of incongruent color words; response latency (ms) and error rate were recorded. A composite Stroop Interference Score was computed as the standardized difference between congruent and incongruent trial response times; larger interference scores indicate

weaker attentional control. The Digit Span Subtest of the WAIS-IV (Wechsler, 2008) measured working memory and sustained attention through forward and backward digit sequence recall.

Critical Thinking

Critical thinking was assessed using Forms D and E of the Watson-Glaser Critical Thinking Appraisal (WGCTA; Watson & Glaser, 1980), a widely validated 80-item instrument comprising five subscales: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. Forms were counterbalanced across pre- and post-test administration to control for form familiarity. The WGCTA has demonstrated robust psychometric properties across educational and professional contexts (Cronbach's α = .81; split-half reliability = .76).

Control Variables

The following control variables were measured at baseline and included as covariates in all analyses: (1) IQ estimated via Raven's Progressive Matrices (18-item short form); (2) years of formal education; (3) socioeconomic status (SES) via parental income bracket and occupation index; (4) daily screen time (device logs + self-report); (5) AI literacy via the 12-item AI Literacy Scale (AILS; Wang et al., 2023; α = .89); (6) introversion-extraversion (Mini-IPIP, 4 items).

3.7 Analytic Strategy

All analyses were conducted in R (version 4.3.1) and Mplus 8.0. Primary analyses employed Analysis of Covariance (ANCOVA) with pre-test scores, IQ, education, and SES as covariates; Difference-in-Differences (DiD) estimation for causal effect inference; multilevel Structural Equation Modeling (SEM) with latent variables for mediation testing (5,000 bootstrap resamples for indirect effects); and moderated regression for AI literacy moderation (Johnson-Neyman technique for floodlight analysis). Intent-to-treat (ITT) and per-protocol (PP) analyses were both conducted; results are reported for ITT with PP in supplementary materials. Effect sizes are reported as Cohen's d for between-group comparisons and standardized regression coefficients (β) for SEM paths. Alpha was set at .05 with Benjamini-Hochberg correction for multiple comparisons.

4. Results

4.1 Compliance and Treatment Fidelity

In the treatment group, mean daily AI tool usage increased from 0.4 hours/day at baseline to 3.1 hours/day at Week 9 (SD = 0.9 hours/day), with ChatGPT used most frequently (62% of logged interactions), followed by Claude (21%), Perplexity (11%), and Gemini (6%). Query log classification revealed that 34% of queries were categorized as Deep Reasoning and Argumentation, 29% as Explanatory Dialogue, 24% as Task Completion, and 13% as Factual Retrieval—indicating that the majority of AI engagement fell within the deeper engagement categories. Social media compliance was high: 88% of treatment participants maintained social media use below the 15-minute/week threshold across the intervention; 12% (n = 18) recorded isolated violations but remained in intent-to-treat analyses.

Control participants maintained stable social media use throughout (baseline: M = 3.3 hrs/day; Week 9: M = 3.4 hrs/day), with no significant change in daily screen time composition.

4.2 Primary Outcome Results

Table 2

Pre-Test, Post-Test, and Follow-Up Means (SD) by Condition and Between-Group Effect Sizes (Cohen's d)

Outcome	Group	Pre-Test M (SD)	Post-Test M (SD)	Follow-Up M (SD)	d (95% CI)
Learning Ability (z)	Treatment	0.02 (0.98)	0.41 (0.91)	—	0.39 [0.18, 0.60]**
	Control	0.01 (1.01)	0.05 (0.99)	—	
Knowledge Retention (%)	Treatment	51.3 (11.2)	—	68.4 (9.8)	0.61 [0.40, 0.82]***
	Control	50.9 (10.8)	—	54.1 (11.4)	
Cognitive Absorption (CAS)	Treatment	3.81 (0.74)	4.62 (0.68)	—	0.48 [0.27, 0.69]***
	Control	3.79 (0.71)	3.94 (0.73)	—	
Stroop Interference (ms)	Treatment	187.4 (42.3)	148.9 (38.6)	—	−0.54 [−0.75, −0.33]***
	Control	185.1 (40.7)	183.6 (41.2)	—	
Digit Span (forward+back)	Treatment	13.2 (2.7)	14.6 (2.6)	—	0.37 [0.16, 0.58]**
	Control	13.4 (2.5)	13.5 (2.6)	—	
Critical Thinking (WGCTA)	Treatment	52.1 (8.4)	56.8 (7.9)	—	0.28 [0.08, 0.49]*
	Control	51.7 (8.6)	53.9 (8.3)	—	

*Note. d values reflect between-group (Treatment vs. Control) effect sizes at post-test (learning, absorption, attention, critical thinking) or follow-up (retention). Stroop interference: lower = better; d reported as negative to indicate treatment improvement. * p < .05. ** p < .01. *** p < .001 (Benjamini-Hochberg corrected).*

Learning Ability (H1)

ANCOVA revealed a significant main effect of condition on post-test learning ability after controlling for pre-test scores, IQ, education, and SES, $F(1, 306) = 14.32$, $p < .001$, $\eta^2 p = .045$. Treatment participants showed a mean standardized learning gain of 0.39 SD units relative to controls (95% CI [0.18, 0.60], d = 0.39). DiD analysis confirmed the net causal treatment effect: Δ(Treatment) – Δ(Control) = 0.43 SD units, $p < .001$. H1 was supported.

Knowledge Retention (H2)

Treatment participants retained a mean of 68.4% of studied material at 30-day follow-up, compared to 54.1% among controls—a difference of 14.3 percentage points ($t(310) = 9.84$, $p < .001$, d = 0.61, 95% CI [0.40, 0.82]). This is the largest observed effect in the study, suggesting that AI-mediated elaborative processing—through dialogue, repeated querying, and self-explanation—substantially enhances the durability of encoded knowledge. H2 was strongly supported.

Sustained Attention (H3)

Stroop interference scores declined significantly in the treatment group (pre: M = 187.4 ms; post: M = 148.9 ms; reduction of 38.5 ms) relative to controls (pre: M = 185.1 ms; post: M = 183.6 ms; reduction of 1.5 ms), yielding a between-group effect of d = 0.54 ($F(1, 306) = 22.18$, $p < .001$, $\eta^2 p = .068$). Digit span composite scores also improved significantly in the treatment group relative to controls (d = 0.37, $p = .002$). These findings indicate meaningful recovery of cognitive inhibition and working memory capacity following social media elimination and AI substitution. H3 was supported.

Cognitive Absorption

Treatment participants reported significantly higher cognitive absorption during AI tool use (post-test CAS M = 4.62, SD = 0.68) compared to control participants' absorption during social media use (M = 3.94, SD = 0.73), $F(1, 306) = 18.07$, $p < .001$, d = 0.48. This finding is noteworthy given that social media is typically described as highly engaging—suggesting that AI interaction may produce deeper and more cognitively involving engagement than passive social media consumption, even though the latter may be experienced as more immediately rewarding.

Critical Thinking (H4)

Overall, the treatment group showed a statistically significant but modest advantage over controls on WGCTA scores at post-test (d = 0.28, $p = .024$). However, this aggregate result masked a theoretically critical interaction with AI literacy. A moderated regression analysis revealed a significant Condition × AI Literacy interaction term ($\beta = 0.31$, $p = .008$, $\Delta R^2 = .026$). Johnson-Neyman floodlight analysis indicated that the treatment advantage on critical thinking was statistically significant only among participants with AI Literacy Scale scores above 29 (on a 0–48 scale), corresponding to the upper 54% of the sample. Among participants with AI literacy scores below 24 (lower 31% of sample), the treatment showed no significant critical thinking benefit (d = 0.09, $p = .41$), and a non-significant trend toward slightly lower critical thinking scores relative to controls. H4a was supported; H4b received partial support.

4.3 Mediation Analysis (H5)

A multilevel SEM model was estimated with AI substitution condition (dummy-coded) predicting post-test learning composite and retention scores, with Stroop interference change (attention recovery) and CAS change (cognitive engagement) entered as mediators. Model fit was excellent: CFI = .971, RMSEA = .048 [90% CI: .031, .065], SRMR = .052, meeting Hu and Bentler's (1999) recommended criteria.

Attention recovery (Stroop interference reduction) significantly mediated the effect of AI substitution on both learning ability (indirect effect: β = .18, 95% bootstrap CI [.09, .29]) and knowledge retention (indirect effect: β = .22, 95% bootstrap CI [.12, .34]). Cognitive engagement (CAS increase) also mediated effects on learning (indirect effect: β = .14, 95% bootstrap CI [.06, .23]) and retention (indirect effect: β = .19, 95% bootstrap CI [.09, .30]). The direct effects of AI substitution on learning and retention remained significant after accounting for mediators, indicating partial mediation. H5 was supported.

Table 3

Structural Equation Model: Standardized Path Coefficients (β) for Direct and Indirect Effects

Path	β	SE	95% CI	p
AI Substitution → Attention Recovery	.44	.06	[.33, .55]	< .001
AI Substitution → Cognitive Engagement	.38	.06	[.27, .49]	< .001
Attention Recovery → Learning	.31	.07	[.18, .44]	< .001
Attention Recovery → Retention	.36	.07	[.23, .49]	< .001
Cognitive Engagement → Learning	.26	.07	[.13, .39]	< .001
Cognitive Engagement → Retention	.29	.07	[.16, .42]	< .001
AI Substitution → Learning (direct)	.19	.07	[.06, .32]	.005
AI Substitution → Retention (direct)	.24	.07	[.11, .37]	.001
AI Substitution → Learning (total indirect)	.32	.06	[.21, .43]	< .001
AI Substitution → Retention (total indirect)	.41	.06	[.30, .52]	< .001

Note. Indirect effects computed via 5,000 bootstrap resamples. Model fit: CFI = .971, RMSEA = .048, SRMR = .052. All covariates (IQ, education, SES, baseline screen time) included but not shown.

4.4 Moderating Effect of AI Literacy

Building on the moderation finding reported for critical thinking, secondary moderation analyses examined whether AI literacy moderated the substitution effects on other cognitive outcomes. AI literacy significantly moderated the effect of AI substitution on learning ability (β_interaction = 0.24, p = .031) and retention (β_interaction = 0.19, p = .048), with stronger treatment effects at higher AI literacy levels. These interactions were smaller in magnitude than the critical thinking moderation and did not alter the

direction of effects for any literacy level above the 10th percentile. Attention recovery showed no significant moderation by AI literacy (β_interaction = 0.08, p = .29), suggesting that attentional benefits of social media elimination are relatively uniform across literacy levels and may operate primarily through the elimination of social media rather than the quality of AI engagement.

4.5 AI Query Depth and Dose-Response Analysis

Among treatment participants, a dose-response analysis examined whether the proportion of Deep Reasoning/Argumentation queries (vs. Factual Retrieval queries) predicted cognitive outcomes above and beyond hours of AI use. Percentage of deep-category queries significantly predicted knowledge retention (β = 0.29, p = .004), critical thinking gains (β = 0.34, p < .001), and cognitive absorption (β = 0.26, p = .009) after controlling for total AI hours, baseline AI literacy, and other covariates. Total AI hours did not significantly predict critical thinking gains (β = 0.11, p = .18) independent of query depth. This pattern supports the theoretical construct of AI engagement quality as a mechanism distinct from simple quantity of AI use.

5. Discussion

5.1 Summary of Findings

This pre-registered RCT provides the first causal evidence that systematically replacing social media use with generative AI tools produces measurable cognitive benefits among young adults over a 60-day intervention. The most robust effects emerged for knowledge retention (d = 0.61), sustained attention recovery as indexed by Stroop interference reduction (d = 0.54), and cognitive absorption (d = 0.48). Learning ability improvements were significant but smaller in magnitude (d = 0.39). Critical thinking effects were heterogeneous, positive overall but concentrated among participants with higher AI literacy, consistent with the moderated hypothesis (H4).

5.2 Attentional Recovery as the Primary Mechanism

The mediation analysis identified attention recovery as the dominant mediating pathway between AI substitution and cognitive outcomes. This finding aligns with Attention Scarcity Theory (Simon, 1971) and with accumulating neuroscientific evidence that social media's notification-driven architecture chronically depletes sustained attentional capacity (Wilmer et al., 2017). The finding that attentional improvement was not moderated by AI literacy—unlike learning and critical thinking effects—suggests that a substantial portion of treatment benefit derives from social media elimination per se, rather than from the specific quality of AI engagement that replaces it. This has a practically important implication: even users who engage with AI tools at a relatively shallow level may benefit attentionally from eliminating social media, while deeper cognitive benefits (retention, critical thinking) require more deliberate, higher-quality AI engagement.

The Stroop interference reduction of 38.5 ms observed in the treatment group is clinically and practically meaningful. Population-level norms indicate that a reduction of this magnitude corresponds to approximately 0.9 standard deviations of developmental improvement in attentional control—a magnitude typically associated with years of maturation (Golden, 1978). While the neurological mechanisms

underlying this behavioral improvement cannot be confirmed without neuroimaging, the pattern is consistent with neuroplasticity models predicting partial restoration of attention-regulatory circuits following removal of chronic attentional disruption.

5.3 Knowledge Retention: The Role of Elaborative Encoding

The largest effect in the study—long-term knowledge retention ($d = 0.61$)—is theoretically interpretable through the lens of Elaborative Interrogation and the Testing Effect (Roediger & Butler, 2011). Generative AI interaction naturally elicits both: users who engage dialogically with AI tools tend to repeatedly reformulate, explain, and interrogate content across multiple queries, a pattern that produces deeper encoding and more retrieval-resistant memory traces than the brief, non-elaborated exposures characteristic of social media information consumption. The dose-response finding that query depth predicts retention independently of query frequency lends direct empirical support to this interpretation.

5.4 The Critical Thinking Moderation: Enhancer or Crutch?

The conditional nature of AI's critical thinking effects—beneficial for high-AI-literacy users, negligible for low-literacy users—is perhaps the study's most theoretically nuanced and practically consequential finding. It directly engages the central intellectual question motivating this research: is generative AI a cognitive enhancer or a cognitive crutch? The present data suggest that the answer is: it depends on how it is used, and user literacy determines how it is used.

High-AI-literacy participants were significantly more likely to use AI for Deep Reasoning queries—generating counterarguments, stress-testing their own positions, requesting evidence evaluation and logical analysis. This pattern produces genuine critical thinking practice: the AI functions as a sophisticated sparring partner that extends rather than replaces the user's own reasoning. Low-AI-literacy participants, by contrast, used AI disproportionately for Factual Retrieval and Task Completion queries—essentially outsourcing information-gathering and decision execution without engaging in the reasoning processes that such activities would otherwise require. For these users, the cognitive substrate of critical thinking received reduced practice, potentially explaining the null or slightly negative trends observed.

This finding has significant implications for the educational technology and AI literacy communities. It suggests that AI's cognitive consequences are not determined by the technology itself but by the pedagogical relationship users develop with it. The implication is not that AI should be restricted but that AI literacy—the metacognitive capacity to use AI as a reasoning partner rather than a reasoning substitute—represents a critical educational objective whose neglect may turn a potential cognitive enhancer into a cognitive dependency.

5.5 Cognitive Absorption and Engagement Quality

The finding that AI tool use produced higher cognitive absorption scores than social media use—despite social media's reputation for addictive engagement—merits careful interpretation. The Cognitive Absorption Scale (Agarwal & Karahanna, 2000) measures not hedonic engagement (experienced pleasure or entertainment) but the depth of cognitive involvement: temporal dissociation, focused immersion, and felt mastery. That AI interaction scores higher on this dimension than social media suggests that, while social media may feel more immediately rewarding (consistent with its dopaminergic variable

reinforcement design), AI interaction produces more cognitively involving engagement. This distinction between hedonic pull and cognitive depth may be important for understanding how these technologies differentially affect the attentional and learning systems.

5.6 Theoretical Contributions

This study advances several theoretical frameworks. It extends Cognitive Load Theory to the domain of AI-human interaction, providing behavioral evidence that dialogic AI engagement reduces extraneous cognitive load and increases germane processing. It empirically tests, for the first time in a substitution RCT, the dual-process prediction that social media's chronic System 1 recruitment impairs System 2 capacity—and that this can be partially reversed within 60 days. It contributes to SDT by demonstrating that AI interaction fulfills competence and autonomy needs in ways that translate into intrinsically motivated deep learning, at least among users who bring sufficient AI literacy to the interaction.

The study also introduces and empirically validates the construct of AI engagement quality—indexed by query depth classification—as a theoretically critical moderator of AI's cognitive consequences. This construct is conceptually distinct from AI usage quantity and suggests a new research agenda focused on the pedagogical and design factors that promote deep versus shallow AI engagement. The moderated mediation findings generate a conceptual model that integrates attention recovery, cognitive engagement quality, and user literacy into a unified causal account of AI substitution's cognitive effects.

5.7 Limitations

Several limitations qualify these findings. First, the Hawthorne Effect—participants' tendency to modify behavior when under observation—poses a threat to ecological validity that cannot be fully eliminated despite the naturalistic design. Treatment participants who know they are being studied for cognitive outcomes may engage more deliberately with AI tools than typical users would, potentially inflating effect sizes. Second, the 60-day duration, while substantially longer than most digital detox studies, may be insufficient to capture neuroplastic changes fully; the attentional gains observed may continue to accrue—or may plateau and reverse—over longer periods. Neuroimaging studies are needed to directly assess neural change. Third, the urban English-speaking sample limits generalizability to other demographic groups, and cultural context may substantially alter both social media use dynamics and AI engagement patterns. Fourth, the classification of AI queries into depth categories relied on a rubric developed for this study; while inter-rater reliability was high (Cohen's $\kappa = .82$), validation in other samples is needed. Fifth, although ITT and PP analyses yielded highly convergent results, compliance monitoring—while multi-method—remains imperfect; some degree of treatment contamination in both directions is likely.

5.8 Conclusions

This study provides the first causal experimental evidence that systematically replacing social media with generative AI tools produces meaningful cognitive benefits for young adults across attention, retention, and learning dimensions over a 60-day period. The benefits are real, but heterogeneous: attentional improvements appear relatively robust across users and appear to be driven substantially by the elimination of social media's chronic attentional disruption; learning and retention gains are amplified by higher-quality AI engagement; and critical thinking benefits are concentrated among users who possess

the AI literacy to use these tools as reasoning partners rather than reasoning substitutes. These findings carry a clear practical conclusion: AI's cognitive potential cannot be realized merely by deploying it—it must be taught, and the pedagogy of AI engagement is as important as the technology itself.

6. Ethical Considerations

The study protocol received full ethical approval from the Institutional Review Board (Protocol #2023-0847). All participants provided written informed consent with explicit disclosure of study purposes, data collection methods (including screen time monitoring and AI query logging), and their right to withdraw without penalty. Informed consent documentation emphasized that AI query logs were anonymized and analyzed only at the category level, not for content surveillance.

Social media abstinence presented genuine psychological risks, given social media's role in social maintenance and emotional regulation for many young adults. The protocol included weekly PHQ-4 wellbeing screening, a dedicated participant support channel staffed by a licensed clinical psychologist consultant, and a pre-specified stopping rule: if more than 10% of treatment participants reported clinically significant distress elevations (PHQ-4 $\geq$ 6), the intervention would pause for IRB review. In the event, 7.4% of treatment participants reported transient distress during Weeks 1–2 of the intervention (associated with social connection concerns); all resolved by Week 4 without clinical intervention. No stopping rule was triggered.

Participants were debriefed at study conclusion with explicit guidance on healthy AI use practices and digital self-regulation resources, given the potential for AI dependency following intensive use. Follow-up contact at 3 and 6 months post-study assessed sustained behavioral changes.

7. Policy and Practical Implications

The findings carry actionable implications across several domains. For universities and educational institutions, the study provides empirical support for incorporating structured AI engagement into curricula not as a supplementary tool but as a primary learning modality, alongside AI literacy training as a mandatory component of first-year programs. Crucially, AI literacy should be framed not merely as technical proficiency but as a metacognitive skill: knowing when to use AI as a reasoning partner versus when independent thought is necessary. The critical thinking moderation finding suggests that deploying AI tools without accompanying literacy instruction may produce no critical thinking benefit—and potentially harm.

For public health policy, the attentional findings join a growing literature indicating that social media's cognitive costs for young adults are measurable and practically significant. A substitution framing—redirecting time from passive social media consumption toward active, dialogic AI engagement—offers a more tractable behavioral intervention target than simple abstinence, given the social and professional functions social media serves. Policymakers might consider evidence-based guidelines analogous to physical activity recommendations: not eliminating sedentary digital behavior but replacing a portion of low-quality (passive social media) screen time with high-quality (active AI engagement) screen time.

For EdTech firms and AI developers, the dose-response finding that query depth—not query frequency—drives cognitive outcomes generates a design imperative: features that promote deep AI engagement (Socratic questioning modules, reflection prompts, argumentation scaffolds, deliberate ignorance nudges that push users to reason before revealing answers) should be prioritized over features that minimize cognitive effort in obtaining AI outputs. The most convenient AI is not the most cognitively beneficial AI.

For employers and workplace productivity, the attentional gains documented here suggest that corporate digital wellbeing initiatives that reduce social media access during focused work periods may benefit from coupling reduction with structured AI tool access, rather than simple restriction, to channel attentional resources toward productive cognitive engagement rather than leaving the attentional vacuum unfilled.

8. Future Research Directions

Several priority directions emerge from this study. First, neuroimaging and EEG-based replication is essential to confirm the neural mechanisms underlying observed behavioral changes. EEG measures of alpha and theta band power during AI interaction versus social media use would directly test attentional engagement predictions; fMRI studies examining prefrontal and default mode network activation during deep versus shallow AI engagement would illuminate the neural correlates of AI engagement quality.

Second, longitudinal studies of one to three years are needed to determine whether attentional and cognitive gains accumulate, stabilize, or reverse over time, and to track AI dependency trajectories as the technology becomes more capable. The present study's 60-day window captures early adaptation but cannot speak to long-term cognitive equilibria.

Third, the study should be replicated with adolescent samples (13–17), given the particular neurological significance of this developmental window for prefrontal maturation and the intensity of social media use in this cohort. Fourth, cross-cultural replication across East Asian, South Asian, Latin American, and Sub-Saharan African samples is essential for establishing boundary conditions; social media use patterns, AI tool availability, and cultural norms around technology and independent reasoning differ substantially across these contexts. Fifth, a three-arm design comparing social media use, generative AI use, and traditional deep reading would clarify whether AI's benefits are unique to its interactive architecture or attributable to any sustained non-social-media intellectual engagement—a comparison with profound implications for digital minimalism advocacy and educational technology investment.

References

1. Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. MIS Quarterly, 24(4), 665–694. https://doi.org/10.2307/3250951

2. Alter, A. (2017). Irresistible: The rise of addictive technology and the business of keeping us hooked. Penguin Press.

3. Andreassen, C. S., Pallesen, S., & Griffiths, M. D. (2016). The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. Addictive Behaviors, 64, 287–293. https://doi.org/10.1016/j.addbeh.2016.03.006

4. Baumgartner, S. E., Weeda, W. D., van der Heijden, L. L., & Huizinga, M. (2014). The relationship between media multitasking and executive function in early adolescents. Journal of Early Adolescence, 34(8), 1120–1144. https://doi.org/10.1177/0272431614527504

5. Biggs, J. B. (1987). Student approaches to learning and studying. Australian Council for Educational Research.

6. Carr, N. (2010). The shallows: What the internet is doing to our brains. W. W. Norton & Company.

7. Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

8. Danaher, J. (2018). Toward an ethics of AI assistants: An initial framework. Philosophy & Technology, 31(4), 629–653. https://doi.org/10.1007/s13347-018-0317-3

9. Davenport, T. H., & Beck, J. C. (2001). The attention economy: Understanding the new currency of business. Harvard Business School Press.

10. DataReportal. (2024). Global social media statistics. https://datareportal.com/social-media-users

11. Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

12. Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

13. Giedd, J. N. (2004). Structural magnetic resonance imaging of the adolescent brain. Annals of the New York Academy of Sciences, 1021(1), 77–85. https://doi.org/10.1196/annals.1308.009

14. Golden, C. J. (1978). Stroop Color and Word Test. Stoelting Company.

15. Haynes, T. (2018). Dopamine, smartphones & you: A battle for your time. Science in the News, Harvard Graduate School of Arts and Sciences. https://sitn.hms.harvard.edu/flash/2018/dopamine-smartphones-battle-time

16. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

17. Hutchins, E. (1995). Cognition in the wild. MIT Press.

18. Jonassen, D. H. (1994). Technology as cognitive tools: Learners as designers. IT Forum Paper, 1, 1–7.

19. Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

20. Lissak, G. (2018). Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. Environmental Research, 164, 149–157. https://doi.org/10.1016/j.envres.2018.01.015

21. Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. British Journal of Educational Psychology, 46(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

22. Merzenich, M. (2013). Soft-wired: How the new science of brain plasticity can change your life. Parnassus Publishing.

23. Moghavvemi, S., Sulaiman, A., Jaafar, N. I., & Kasem, N. (2018). Social media as a complementary learning tool for teaching and learning: The case of YouTube. International Journal of Management Education, 16(1), 37–42. https://doi.org/10.1016/j.ijme.2017.12.001

24. Mollick, E. R., & Mollick, L. (2023). Assigning AI: Seven approaches for students, with prompts. SSRN Working Paper. https://doi.org/10.2139/ssrn.4475995

25. Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

26. Nye, B. D., Graesser, A. C., & Hu, X. (2023). AutoTutor and family: A review of 17 years of natural language tutoring. International Journal of Artificial Intelligence in Education, 33(1), 89–105. https://doi.org/10.1007/s40593-022-00308-7

27. Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

28. Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences, 15(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

29. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

30. Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), Computers, communication, and the public interest (pp. 37–72). Johns Hopkins Press.

31. Statista. (2024). Social media usage by age group worldwide. https://www.statista.com/statistics/471370

32. Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. Developmental Review, 28(1), 78–106. https://doi.org/10.1016/j.dr.2007.08.002

33. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

34. Thornton, B., Faires, A., Robbins, M., & Rollins, E. (2014). The mere presence of a cell phone may be distracting: Implications for attention and task performance. Social Psychology, 45(6), 479–488. https://doi.org/10.1027/1864-9335/a000216

35. Tromholt, M. (2016). The Facebook experiment: Quitting Facebook leads to higher levels of well-being. Cyberpsychology, Behavior, and Social Networking, 19(11), 661–666. https://doi.org/10.1089/cyber.2016.0259

36. Tulving, E. (1983). Elements of episodic memory. Oxford University Press.

37. Twenge, J. M., & Campbell, W. K. (2019). Media use is linked to lower psychological well-being: Evidence from three datasets. Psychiatric Quarterly, 90(2), 311–331. https://doi.org/10.1007/s11126-019-09630-7

38. Valkenburg, P. M., Meier, A., & Beyens, I. (2022). Social media use and its impact on adolescent mental health: An umbrella review of the evidence. Current Opinion in Psychology, 44, 58–68. https://doi.org/10.1016/j.copsyc.2021.08.017

39. Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow-up measurements. BMJ, 323(7321), 1123–1124. https://doi.org/10.1136/bmj.323.7321.1123

40. Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

41. Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using AI literacy: A scale development and validation study. Behaviour & Information Technology, 42(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2079266

42. Watson, G., & Glaser, E. M. (1980). Watson-Glaser Critical Thinking Appraisal. Psychological Corporation.

43. Wechsler, D. (2008). Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). Pearson.

44. Wilmer, H. H., Sherman, L. E., & Chein, J. M. (2017). Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. Frontiers in Psychology, 8, 605. https://doi.org/10.3389/fpsyg.2017.00605

45. Wolf, M. (2018). Reader, come home: The reading brain in a digital world. Harper.