Cognitive Debt: Neural and Behavioral Evidence of AI-Induced Thinking Atrophy

Anonymous authors
Paper under double-blind review

Abstract

Over half of U.S. adults now use generative AI, faster adoption than personal computers or the internet saw at comparable stages. We synthesize emerging evidence on what this cognitive outsourcing actually does to human minds. The picture is troubling: AI use correlates with reduced neural engagement, weaker critical thinking, and systematic overestimation of AI's benefits. Consider the perception-reality gap: developers using AI complete tasks 19% slower while believing they're 20% faster. We introduce the concept of "cognitive debt" from neurophysiological evidence showing persistent underengagement after AI use, and identify three trajectory-shaping factors: the absence of constructive friction in current interfaces, disappearing entry-level cognitive work, and the emergence of AI-only social ecosystems. Drawing on the Technology-Induced Cognitive Diminishment framework and the "System 0" proposal, we analyze three plausible futures: human-AI symbiosis, cognitive stratification, and progressive dependency. Habits of mind are forming now that may prove difficult to reverse.

1 Introduction

This went from speculative concern to measurable phenomenon faster than anyone expected. As of mid-2025, over half of U.S. adults use generative AI (Pew Research Center, 2025), adoption outpacing personal computers and the internet at comparable stages. What's different this time? Calculators offload arithmetic. GPS offloads navigation. LLMs engage with open-ended reasoning, writing, and analysis: the processes that define human intellectual work.

The stakes climb higher as we approach AGI. Today's AI tools assist human cognition; AGI-level systems may render human cognition unnecessary for most intellectual tasks. The question shifts from "how do we use AI effectively?" to "what cognitive capacities must humans retain to maintain meaningful agency in a post-AGI world?" Early evidence points to a troubling pattern: AI use correlates with **reduced critical thinking** and **decreased neural engagement**, even as users report feeling sharper (Gerlich, 2025; Kosmyna et al., 2025). This perception-reality gap could accelerate dependency without anyone noticing.

We synthesize the emerging empirical evidence, examine theoretical frameworks for human-AI cognitive integration, and map out three plausible futures. The P-AGI research community faces an urgent challenge: how do we maintain meaningful human cognitive engagement as AI capabilities scale?

2 Empirical Evidence of Cognitive Outsourcing

2.1 Neural Evidence: The "Cognitive Debt" Finding

An MIT Media Lab study used EEG to track brain activity during essay writing (Kosmyna et al., 2025). Fifty-four participants, four months, three conditions: brain-only, search engine, or Chat-GPT. LLM users showed **systematically weaker neural connectivity** in alpha and beta bands, the frequencies tied to focused attention and complex cognition.

The striking finding was "cognitive debt": participants who switched from LLM assistance back to brain-only writing showed **persistent underengagement**. Their brains seemed habituated to reduced effort. These users also reported feeling less ownership of their essays and couldn't accurately recall what they'd written. The sample is small (n=54) and selection effects may be at play, so caution is warranted. Still, this offers early neurophysiological evidence that AI assistance may alter how we engage during intellectual work.

2.2 CRITICAL THINKING AND THE PERCEPTION-REALITY GAP

A cross-sectional study of 666 participants found a **strong negative correlation** ($r = -0.75$) between cognitive offloading and critical thinking abilities, with AI use driving offloading ($r = 0.72$) (Gerlich, 2025). Younger participants showed higher AI dependence and lower critical thinking scores; higher education seemed protective. The design can't establish causation (maybe weaker critical thinkers just gravitate toward AI tools), but that correlation is hard to ignore.

The perception-reality gap shows up starkly in productivity research. One randomized trial found GitHub Copilot users finished tasks 55.8% faster on greenfield development tasks (Peng et al., 2023). But here's the twist: the METR study found experienced developers working in their *own* codebases were **19% slower** with AI assistance, while perceiving themselves 20% faster (Becker et al., 2025). That's a 39 percentage-point gap between perception and reality. At least in some contexts, we may be systematically misjudging what AI actually contributes.

Fernandes et al. (2026) provide evidence that AI may eliminate the typical Dunning-Kruger pattern. *All* AI users, regardless of baseline skill, systematically overestimate their performance. Worse, higher AI literacy correlates with *greater* overconfidence. Technical sophistication doesn't protect against metacognitive miscalibration; it may amplify it.

2.3 HISTORICAL PRECEDENTS AND DESKILLING

Humans have always offloaded cognitive work. The "Google Effect" showed that people remember less when they expect information to stay accessible, though they remember better *where* to find it (Sparrow et al., 2011). GPS research makes the case more strongly: lifetime GPS use correlates negatively with spatial memory ($r = -0.20$ to -0.22), and longitudinal follow-up confirmed GPS use *caused* steeper declines in hippocampal-dependent memory (Dahmani & Bohbot, 2020).

Aviation offers a warning. Operators over-trust automated systems, leading to monitoring failures (Parasuraman & Manzey, 2010). Pilots show "skill erosion" from automated flights, not just manual skills but situational awareness and cognitive agility. Air France 447 showed how proficiency erodes through over-reliance on automation. Medical researchers warn that AI dependence risks deskilling clinicians by reducing the cognitive engagement needed to maintain diagnostic expertise (El Tarhouny & Farghaly, 2026).

Then there's the "missing ladder rung" problem. When AI handles entry-level cognitive work, newcomers don't develop professional intuition. Bureau of Labor Statistics data show a **27.5% decline** in computer programmer employment since 2023 (Bureau of Labor Statistics, 2025). UK tech graduate roles fell 46% in 2024, with projections for another 53% decline by 2026. Gartner analyst Daryl Plummer put it bluntly at their October 2025 symposium: "AI is stealing your skills" (Plummer, 2025).

3 THEORETICAL FRAMEWORKS

3.1 SYSTEM 0: AI AS COGNITIVE PREPROCESSOR

Chiriatti et al. (2024) extend Kahneman's dual-process theory with "System 0," AI as a cognitive preprocessor shaping information *before* System 1 (intuitive) or System 2 (analytical) thinking kicks in. Unlike our internal systems, System 0 operates outside the mind, embedded in our technological environment.

This framing highlights a subtle danger. If AI preprocesses our cognitive inputs, it constrains the solution space before conscious deliberation even begins. The risk isn't that AI replaces human thinking. It's that AI shapes what we think *about*.

3.2 EXTENDED MIND VS. COGNITIVE EXTRACTION

Clark and Chalmers' Extended Mind thesis argues that external processes functioning like internal cognition *are* part of the cognitive system (Clark & Chalmers, 1998). By this logic, AI literally becomes part of human minds. But a counter-framework is emerging: the "extracted mind" thesis. Here, AI doesn't scaffold cognition; it "captures, mimics, and then replaces" cognitive skills.

The Technology-Induced Cognitive Diminishment (TICD) framework (Fasoli et al., 2025) systematically distinguishes enhancement from displacement. The question sharpens: when does cognitive extension become cognitive extraction?

3.3 EPISTEMIC AUTONOMY AND TECHNOMORAL VIRTUES

Carter (2022) argues virtuous epistemic autonomy means relying on external resources "to an appropriate extent." But what's appropriate for AI? That remains unresolved. Vallor (2016) warns of "moral deskilling" through habit-forming tool use. The character traits that enable flourishing in technological contexts must now include wisdom about AI partnership.

The challenge isn't binary. It's calibrational. Can we develop the metacognitive capacity to know when AI helps versus hinders? And the self-regulation to act on that knowledge?

3.4 THE CASE FOR OPTIMISM

Fair enough. Consider the counter-arguments. Clark (2025) argues in *Nature Communications* that humans have always been "hybrid thinking systems." Worrying that AI makes us "dumber" misses something fundamental about cognition. He points to Go: after AI strategies emerged, human players showed *increased* novelty, exploring previously neglected corners of strategy space rather than just copying AI moves.

Cognitive offloading does have real benefits. It conserves resources for higher-order tasks, enables problem-solving beyond biological limits, and reduces cognitive load during stress. The relationship between AI use and cognitive effects may be nonlinear; moderate use might not impair critical thinking much.

Our response is calibrational, not prohibitionist. We're not worried about AI use per se. We're worried about (1) the perception-reality gap that prevents accurate assessment of AI's contribution, (2) the absence of "constructive friction" in current interfaces, and (3) adoption outpacing our understanding of long-term effects. Clark himself emphasizes "extended cognitive hygiene," critically evaluating what we incorporate into our minds. The empirical evidence suggests current AI interactions may fail to meet even his own standards.

4 THREE PLAUSIBLE FUTURES

4.1 HUMAN-AI SYMBIOSIS

The optimistic scenario realizes Licklider's 1960 vision: humans and computers as collaborative partners, each contributing what the other can't. Meta-analysis finds human-AI teams show **positive synergy in content creation and problem formulation**, but negative synergy in judgment tasks. The path forward: augment formulation, don't automate judgment.

This future demands "AI fluency": orchestrating AI systems while maintaining independent cognitive capacity. Educational systems must build foundational skills *before* introducing AI scaffolding. The generation now in school will either be the last to develop these skills independently or the first to never need them.

4.2 COGNITIVE STRATIFICATION

The Oxford Martin AI Governance Initiative introduces "agentic inequality," the power disparities arising from differential access to AI agents (Sharp et al., 2025). This isn't like previous digital divides. AI agents act as autonomous delegates, creating power asymmetries through scalable goal delegation.

Three dimensions matter: **availability**, **quality**, and **quantity** of agent access. Oxford's research found that "sophisticated corporate agents can consistently exploit less capable consumer agents." The stratification extends to AI-mediated interactions: your agent negotiates with their agent, and the outcome depends on whose is better.

Moltbook (a social network exclusively for AI agents, with over 1.5 million autonomous participants as of January 2026) shows cognitive handoff at scale (Holtz, 2026). AI agents now operate in their own social ecosystems, exhibiting emergent behaviors humans didn't design. Who retains meaningful cognitive agency? The question is no longer hypothetical.

4.3 PROGRESSIVE DEPENDENCY

Call it the "boiling frog" scenario. Incremental AI conveniences erode independent cognitive capacity until a triggering event reveals irreversible fragility. Gartner predicts that by 2026, 50% of organizations will require "AI-free" skills assessments due to critical thinking atrophy (Gartner, Inc., 2025).

The OECD warns that "offloading cognitive tasks to general-purpose chatbots creates risks of metacognitive laziness and disengagement that may deter skill acquisition in the long run" (OECD, 2026). Institutions are responding: the EU AI Act's human oversight requirements take effect August 2026. Four U.S. states have enacted neural data protection laws; the proposed MIND Act would classify "thoughts, emotions, and decision-making patterns" as protected information.

Aviation precedents, the perception-reality gap, the missing ladder rung: all suggest this pathway is the default without deliberate intervention.

5 RELATED WORK

Cognitive Offloading and Extended Mind. The Extended Mind thesis (Clark & Chalmers, 1998) and its application to generative AI (Clark, 2025) argue humans have always been "hybrid thinking systems." However, recent work on the "extracted mind" challenges whether AI scaffolds or displaces cognition.

Technology-Induced Skill Change. Research on GPS and spatial memory (Dahmani & Bohbot, 2020), the Google Effect (Sparrow et al., 2011), and automation bias (Parasuraman & Manzey, 2010) establishes precedents. The TICD framework (Fasoli et al., 2025) provides systematic criteria for distinguishing enhancement from displacement.

AI and Critical Thinking. Emerging work correlates AI use with reduced critical thinking (Gerlich, 2025), decreased neural engagement (Kosmyna et al., 2025), and perception-reality gaps (Becker et al., 2025; Fernandes et al., 2026). Our contribution synthesizes these threads, focusing on the metacognitive miscalibration that makes cognitive outsourcing self-reinforcing.

6 RESEARCH IMPERATIVES AND CONCLUSION

Humans are already outsourcing significant cognitive work to AI, with measurable effects on neural engagement, critical thinking, and metacognitive calibration. Urgent questions remain: can "constructive friction" maintain engagement alongside AI assistance? Is there a critical period for skill acquisition before AI scaffolding is appropriate? What interventions restore cognitive capacities after dependency sets in?

For the P-AGI research community, maintaining meaningful human cognitive engagement isn't just a design preference. It's a prerequisite for the scalable oversight that safe AGI deployment requires. Habits of mind are forming now that may prove difficult to reverse.

REFERENCES

Joel Becker, Nate Rush, Elizabeth Barnes, and David Rein. Measuring the impact of early-2025 AI on experienced open-source developer productivity. *arXiv preprint*, 2025. METR.

Bureau of Labor Statistics. Current population survey: Computer programmers employment. *Bureau of Labor Statistics, U.S. Department of Labor*, 2025. Data reported in Fortune, March 17, 2025.

J. Adam Carter. *Autonomous Knowledge: Radical Enhancement, Autonomy, and the Future of Knowing*. Oxford University Press, Oxford, 2022. ISBN 9780192846921.

Massimo Chiriatti, Marianna Ganapini, Enrico Panai, Mario Ubiali, and Giuseppe Riva. The case for human–AI interaction as system 0 thinking. *Nature Human Behaviour*, 8:1829–1830, 2024. doi: 10.1038/s41562-024-01995-5.

Andy Clark. Extending minds with generative AI. *Nature Communications*, 16:4627, 2025. doi: 10.1038/s41467-025-59906-9.

Andy Clark and David Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998. doi: 10.1093/analys/58.1.7.

Louisa Dahmani and Véronique D. Bohbot. Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, 10:6310, 2020. doi: 10.1038/s41598-020-62877-0.

Shereen El Tarhouny and Amira Farghaly. The deskilling dilemma: Brain over automation in medical practice. *Frontiers in Medicine*, 13:1765692, 2026. doi: 10.3389/fmed.2026.1765692.

Marco Fasoli, Guido Cassinadri, and Marcello Ienca. The dark side of cognitive enhancement: A framework for the technologically induced cognitive diminishment. *Journal of Cognitive Enhancement*, 9:346–359, 2025. doi: 10.1007/s41465-025-00331-7.

Daniela Fernandes, Steeven Villa, Salla Nicholls, Otso Haavisto, Daniel Buschek, Albrecht Schmidt, Thomas Kosch, Chenxinran Shen, and Robin Welsch. AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior*, 175: 108779, 2026. doi: 10.1016/j.chb.2025.108779. Aalto University study.

Gartner, Inc. Gartner's top strategic predictions for 2026 and beyond. Technical report, Gartner, 2025. URL https://www.gartner.com/en/articles/strategic-predictions-for-2026.

Michael Gerlich. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1):6, 2025. doi: 10.3390/soc15010006.

David Holtz. The anatomy of the Moltbook social graph. Columbia Business School Working Paper, 2026. Preliminary draft analyzing the AI-only social network.

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 preprint*, 2025. MIT Media Lab.

OECD. OECD digital education outlook 2026. Technical report, OECD Publishing, Paris, 2026.

Raja Parasuraman and Dietrich H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3):381–410, 2010. doi: 10.1177/0018720810376055.

Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint*, 2023.

Pew Research Center. How Americans view AI and its impact on human abilities, society. Technical report, Pew Research Center, September 2025. URL https://www.pewresearch.org/science/2025/09/17/how-americans-view-ai-and-its-impact-on-people-and-society/.

Daryl Plummer. AI is stealing your skills. Gartner IT Symposium/Xpo Keynote, October 2025. Orlando, Florida.

Matthew Sharp, Omer Bilgin, Iason Gabriel, and Lewis Hammond. Agentic inequality. Research report, Oxford Martin School, AI Governance Initiative, October 2025. URL https://www.oxfordmartin.ox.ac.uk/publications/agentic-inequality.

Betsy Sparrow, Jenny Liu, and Daniel M. Wegner. Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043):776–778, 2011. doi: 10.1126/science.1207745.

Shannon Vallor. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press, New York, 2016. ISBN 9780190498511.

A APPENDIX

LLMs were used for drafting assistance during the preparation of this paper. The authors take full responsibility for all content.