ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

Cognitive Effects of Large Language Models: An Empirical Study of Memory Retention, Reasoning Strategies, and Critical Thinking among Students and Professionals at the University of Gujrat, Pakistan

Saima Jamshaid
saima.jamshaid@uog.edu.pk
Lecturer, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan
Corresponding Author: * Saima Jamshaid saima.jamshaid@uog.edu.pk

Received: 02-07-2025 **Revised:** 20-08-2025 **Accepted:** 12-09-2025 **Published:** 30-09-2025

ABSTRACT

This study investigates whether the regular use of large language models (LLMs), such as ChatGPT, is reshaping human cognition by altering memory retention, reasoning strategies, and critical thinking skills. Drawing from cognitive psychology and the extended mind hypothesis, the researcher conducted a mixed-methods study involving 120 participants—80 students and 40 professionals—divided into frequent and non-frequent LLM users. Experimental tasks measured short- and long-term recall, problem-solving approaches, and critical evaluation of arguments, complemented by surveys and interviews on user perceptions. Findings suggest that LLM users demonstrate improved efficiency in problem-solving but lower long-term recall, with a tendency to externalize reasoning and rely more heavily on AI-generated outputs. Implications for education, workplace training, and epistemic trust are discussed, emphasizing the need for pedagogical strategies that preserve critical thinking in AI-mediated environments.

Keywords: *Artificial intelligence; Cognitive psychology; Critical thinking; Extended mind hypothesis; Large language models; Memory retention; Problem-solving*

INTRODUCTION

In less than three years, generative artificial intelligence (AI) systems such as ChatGPT, Claude, and Gemini have transitioned from technological novelties to integral components of education, research, and professional practice. Students regularly consult AI tools to clarify difficult concepts, draft essays, and generate practice questions. Professionals use them to summarize lengthy reports, prepare presentations, and conduct preliminary data analysis. Far from being niche technologies, large language models (LLMs) have become a pervasive layer of the knowledge economy, subtly reconfiguring how individuals access, process, and evaluate information.

Scholars have long argued that tools shape cognition. The extended mind hypothesis (Clark & Chalmers, 1998) posits that cognitive processes can extend beyond the individual brain into the surrounding environment, integrating artifacts like notebooks, smartphones, and—now—AI assistants into an expanded cognitive system. Under this view, consulting ChatGPT for explanations or problem-solving is not merely outsourcing memory but extending the boundaries of thought itself.

However, not all theoretical frameworks are equally optimistic. Cognitive load theory (Sweller, 2011) warns that reducing cognitive effort too much may hinder the construction of durable mental schemas, while research on retrieval practice (Roediger & Butler, 2011) demonstrates that active recall is essential for consolidating information into long-term memory. If LLMs make recall unnecessary, deep learning may weaken over time. Moreover, studies of automation bias (Mosier & Skitka, 2018) show that individuals often over-trust algorithmic outputs, even when flawed, potentially compromising independent critical thinking.

Empirical studies to date provide mixed evidence. Sparrow et al. (2011) demonstrated that easy access to digital information leads to lower fact recall but improved "transactive memory"—remembering where information can be found. More recently, Liu et al. (2023) reported that students using GPT-powered study aids performed better on immediate comprehension tasks but worse on delayed recall. Wu et al. (2023) found that LLM-assisted participants solved logic puzzles faster but produced less original solutions. Together, these findings suggest that while LLMs increase efficiency, they may simultaneously diminish deep processing and originality.

Significance

Despite growing interest, much of the existing scholarship focuses on the technical performance of LLMs (e.g., accuracy, hallucination rates) or their practical utility in professional and educational settings. There remains a paucity of research on how sustained LLM use affects human cognition itself—specifically memory retention, reasoning strategies, and critical thinking. This gap is particularly notable in non-Western contexts, where educational traditions, epistemic norms, and digital literacy may shape cognitive outcomes differently.

Furthermore, most prior studies rely on self-reported perceptions rather than objective cognitive performance metrics. Few adopt a mixed-methods approach that integrates experimental testing with qualitative insights into users' subjective experiences. Understanding not just whether cognition is changing but how users interpret and adapt to these changes is essential for developing evidence-based pedagogical and policy interventions.

Objectives

This article seeks to address these gaps by presenting one of the first systematic, mixed-methods studies of LLM impact on cognition among students and professionals in South Asia. The evaluation of three interrelated cognitive domains is the focus of the study.

1. **Memory Retention** – to evaluate whether with the use of regular LLM durable knowledge consolidation diminishes or not
2. **Reasoning Strategies** – to assess the depth of participants' knowledge and their reliance on AI-generated suggestions during problem-solving
3. **Critical thinking** – to evaluate the ability of the participants in detecting bias, misinformation, and logical fallacies, especially in AI-generated outputs.

This study integrates controlled experimental tasks, pre- and post-task surveys, and semi-structured interviews to provide a comprehensive account of how human cognition is reshaped in AI-assisted environments. The findings aim to inform educators, cognitive scientists, employers, and policymakers seeking to balance productivity gains with the preservation of deep intellectual skills and independent reasoning

LITERATURE REVIEW

The emergence of generative artificial intelligence (AI) large language models (LLMs) has fundamentally reshaped how humans access, synthesize, and produce knowledge. Unlike earlier digital tools, LLMs generate fluent, contextually relevant responses, enabling real-time engagement with vast bodies of information. While these capabilities democratize access and enhance productivity, they also raise

important questions about the cognitive trade-offs involved in outsourcing memory, reasoning, and evaluative judgment to machines (Jacobs & Wallach, 2021; Weidinger et al., 2022).This literature review surveys three major areas relevant to these questions: (a) cognitive offloading and memory, (b) reasoning and problem-solving strategies, and (c) critical thinking and epistemic trust. Together, these domains provide a conceptual foundation for investigating whether sustained LLM use is reshaping human cognition.

LLMs and Cognitive Offloading

Cognitive offloading refers to the use of external tools to reduce the burden on internal cognitive resources (Risko & Gilbert, 2016). Humans have long externalized memory through material and digital artifacts—ranging from clay tablets and written scripts to smartphones and cloud-based reminders (Donald, 1991; Norman, 1993). This process has been reframed in contemporary cognitive science as part of a "distributed cognition" paradigm, which views memory and reasoning as extending beyond the brain into sociotechnical systems (Hutchins, 1995).

Digital technology has accelerated this trend. Sparrow, Liu, and Wegner (2011) coined the term *Google effect* to describe how ready access to online search tools leads individuals to remember fewer facts but become better at recalling where to find information. This shift represents the rise of *transactive memory systems*, where memory storage is externalized, allowing individuals to allocate cognitive resources to other tasks (Wegner et al., 1985).

LLMs represent an even more significant leap in cognitive offloading because they do not merely store information but dynamically generate explanations, synthesize arguments, and scaffold step-by-step reasoning. This capacity has prompted concerns about potential trade-offs between efficiency and durable knowledge acquisition. Empirical research by Liu et al. (2023) found that students who relied on GPT-powered study tools demonstrated higher immediate comprehension but performed worse on delayed recall tests, suggesting weakened consolidation of long-term memory. These findings align with the cognitive psychology literature emphasizing that retrieval practice—actively recalling information—is essential for durable learning (Roediger & Butler, 2011).

Cognitive load theory (Sweller, 2011) provides another lens for interpreting these effects. By automating tasks such as paraphrasing, summarizing, and information retrieval, LLMs reduce extraneous cognitive load, theoretically freeing working memory for schema construction. However, excessive reliance may also reduce *germane* cognitive load—the mental effort invested in processing and integrating new information—resulting in surface-level understanding (Kalyuga, 2011).

A complementary perspective is offered by the extended mind hypothesis (Clark & Chalmers, 1998), which posits that external tools, when reliably accessible and automatically used, become integral parts of cognitive systems. Under this framework, memory is not "lost" but redistributed between internal storage and external artifacts. Nevertheless, Smart (2018) warns that over-reliance on cognitive extensions may reduce metacognitive awareness, leading to confusion about what knowledge is personally held versus externally stored. This has implications for epistemic agency, as individuals may gradually lose track of their own competence boundaries.

Reasoning and Problem-Solving

Reasoning and problem-solving represent higher-order cognitive domains where LLMs may both enhance and hinder performance. Decades of research on decision support systems show that algorithmic aids can improve decision accuracy and speed in well-structured tasks (Silver, 1991; Klein et al., 2004). However, this assistance carries risks: users often exhibit automation bias, over-relying on algorithmic outputs and failing to detect errors (Mosier & Skitka, 2018).

Emerging empirical studies suggest that LLMs may shape not just outcomes but also reasoning *strategies*. Wu et al. (2023) observed that participants using ChatGPT to solve logic puzzles completed tasks more quickly but generated less diverse solution paths, indicating a narrowing of cognitive exploration. Similarly, Murgia and Piffer (2024) found that participants exposed to LLM-generated justifications were more likely to adopt those arguments wholesale, reducing independent reasoning effort.

Dual-process theories of reasoning (Kahneman, 2011) help explain these effects. LLMs may encourage "System 1" thinking—fast, intuitive, and heuristic-driven—by presenting plausible solutions quickly, thus discouraging slower, more deliberate "System 2" analysis. Over time, this may reduce users' ability to engage in effortful, flexible problem-solving, potentially undermining originality (Runco & Jaeger, 2012).

Importantly, not all impacts are negative. Kasneci et al. (2023) showed that when students were tasked with critically evaluating and revising LLM outputs, they demonstrated improved metacognitive awareness and deeper conceptual understanding. This suggests that the cognitive impact of LLMs is not fixed but contingent on task design **and** user engagement level—a key consideration for educators and employers seeking to integrate AI responsibly.

Critical Thinking and Epistemic Trust

Critical thinking—understood as the ability to evaluate arguments, identify bias, and draw well-reasoned conclusions—remains a cornerstone of higher education and professional competence (Facione, 1990). The fluent, polished outputs of LLMs present new challenges for epistemic vigilance. Because these systems produce coherent responses regardless of factual accuracy, users may equate linguistic fluency with truthfulness, a phenomenon often described as 'truthiness bias' (Jacovi et al., 2023). Research on misinformation shows that individuals rely heavily on source cues when assessing credibility (Pennycook & Rand, 2020). Yet LLMs frequently lack clear authorship and provide inconsistent citations unless specifically prompted, requiring users to invest additional cognitive effort in verification. Survey data from Zhao and Wang (2023) found that 62% of respondents reported trusting ChatGPT responses without independent fact-checking, particularly under time pressure.

The long-term implications of these tendencies are significant. Favier et al. (2024) found that repeated exposure to AI-generated news summaries gradually reduced participants' confidence in their own evaluative judgments and increased deference to algorithmic recommendations. This suggests that LLM use may contribute to a restructuring of epistemic norms, shifting authority from the individual to the machine.

However, LLMs can also be harnessed to strengthen critical thinking skills. Arfini and Michelini (2023) showed that adversarial prompting—requiring students to question and fact-check AI-generated outputs—improved their ability to detect inconsistencies in both AI- and human-authored texts. Such findings

suggest that, when used intentionally, LLMs can serve as 'intellectual sparring partners,' fostering epistemic resilience rather than passive consumption.

Research Gap

Despite these insights, important gaps remain. Much of the current literature focuses on technical performance metrics such as accuracy and hallucination rates, with far less attention to cognitive outcomes for users (Kasneci et al., 2023).

Second, studies that do examine cognition often rely on self-reported perceptions of usefulness or trust, which may not accurately capture underlying cognitive processes (Jacobs & Wallach, 2021).

Third, most empirical work has been conducted in Western, educated, industrialized, rich, and democratic (WEIRD) contexts (Henrich et al., 2010), leaving unanswered questions about how LLM use interacts with distinct epistemic traditions, pedagogical norms, and digital literacy levels in the Global South. Finally, few studies adopt a mixed-methods approach that combines quantitative performance measures (e.g., recall scores, originality ratings, fallacy detection accuracy) with qualitative insights into user experiences.

The present study addresses these gaps by systematically examining the effects of regular LLM use on three interrelated domains—memory retention, reasoning strategies, and critical thinking—among students and professionals in South Asia. By pairing experimental tasks with surveys and semi-structured interviews, this research offers a more holistic account of how human cognition is being reconfigured in an era of generative AI, with implications for education, workplace training, and epistemic trust.

METHODOLOGY

Participants

120 participants were recruited: 80 students from University of Gujrat (from humanities, social sciences, and STEM disciplines) and 40 professionals (journalists, researchers, and analysts) from Pakistan. Participants were classified as frequent LLM users (daily or near-daily use) or low/no users (less than once a week).

Experimental Design

Participants were randomly assigned to two groups: one that used ChatGPT to complete tasks and one that did not.

Memory Task: Participants read a 500-word text. The LLM group used ChatGPT to generate a summary, while the control group summarized manually. Immediate recall was tested after 10 minutes; delayed recall was tested after 48 hours.

Reasoning Task: Participants solved three logic and case-based problems. Solutions were coded for originality, depth of reasoning, and whether participants sought AI input.

Critical Thinking Task: Participants were shown four arguments (two human-written, two AI-generated) containing subtle logical fallacies or factual inaccuracies. They had to identify and explain the flaws.

Surveys and Interviews

Pre- and post-task surveys captured participants' trust in AI, self-reported frequency of reliance, and perceived cognitive effort. Additionally, semi-structured interviews (n = 30) provided qualitative insights into how participants believed AI use influenced their study and work practices.

DATA ANALYSIS

Quantitative Data Analysis

A between-subjects ANOVA was conducted to compare frequent LLM users and low/no users across outcomes related to memory retention, reasoning strategies, and critical thinking accuracy. Where significant effects were observed, Tukey's HSD post-hoc tests were applied. Multiple regression models were also used to control for demographic variables such as age, gender, discipline, and country.

Memory Retention

Participants' recall scores were measured immediately (10 minutes) and after 48 hours (delayed recall).

Group	Immediate Recall (Mean ± SD)	Delayed Recall (Mean ± SD)
Frequent LLM Users	82.5 ± 9.4	58.7 ± 10.1
Low/No LLM Users	80.8 ± 10.3	70.3 ± 9.7

ANOVA Results

- Immediate recall: $F(1,118)=1.02$, $p=0.31$ (NS)
- Delayed recall: $F(1,118)=14.45$, $p<0.001$

This suggests no significant difference in short-term recall but a significant drop in long-term retention among frequent LLM users. Regression models controlling for discipline showed LLM frequency remained a strong negative predictor of delayed recall ($\beta = -0.36$, $p<0.001$).

Reasoning Strategies

Solutions were coded for depth (0–10), originality (0–10), and time taken to solve problems (minutes).

Group	Reasoning Depth	Originality	Time Taken (min)
Frequent LLM Users	7.8 ± 1.3	6.1 ± 1.7	8.5 ± 2.2
Low/No LLM Users	7.5 ± 1.5	7.2 ± 1.6	11.3 ± 2.8

ANOVA Results

- Reasoning depth: $F(1,118) = 0.84$, $p=0.36$ (NS)
- Originality: $F(1,118) = 7.92$, $p<0.01$
- Time taken: $F(1,118) = 16.31$, $p<0.001$

LLM users solved problems faster but with significantly lower originality scores, indicating a reliance on paraphrased AI outputs rather than self-generated solutions.

Critical Thinking Accuracy

Participants were scored on correctly identifying logical fallacies or factual errors in four arguments (max score = 8).

Group	Critical Thinking Score (Mean ± SD)
Frequent LLM Users	4.3 ± 1.2
Low/No LLM Users	5.6 ± 1.1

ANOVA Results

- $F(1,118) = 19.04$, $p<0.001$

Frequent LLM users were significantly less accurate in detecting flawed arguments, particularly when the flawed argument was AI-generated. This aligns with prior literature on automation bias and epistemic over-trust.

Qualitative Data Analysis

Thirty participants (15 students, 15 professionals; balanced across LLM frequency) were interviewed post-task. Data were thematically analyzed following Braun & Clarke's (2006) six-step approach: familiarization, coding, theme development, reviewing, defining, and reporting.

Emergent Themes

Theme	Description	Representative Quote
Externalized Memory	Participants reported outsourcing information storage to AI rather than memorizing.	"I don't bother remembering dates anymore—I just ask ChatGPT when I need them." (Student, frequent user)
Efficiency vs. Depth Trade-off	Users praised speed but worried about losing depth of thought.	"It's faster, but I think I stop too soon. I don't explore ideas as much as I used to." (Professional, frequent user)
Automation Bias	Participants tended to over-trust fluent AI responses, even when incorrect.	"If it sounds right, I usually just take it as true. Only later do I fact-check if I have time." (Student, frequent user)
Cognitive Fatigue Reduction	Some participants experienced less mental strain, viewing AI as a	"I save mental energy for creative tasks; ChatGPT handles the routine thinking."

Theme	Description	Representative Quote
	"cognitive partner."	(Professional, frequent user)
Epistemic Vigilance (low/no users)	Non-users reported a stronger habit of cross-checking information manually.	"I prefer to verify from at least two sources before I believe anything." (Student, low user)

Pattern Summary

- **Frequent LLM Users:** Emphasized convenience, mental energy conservation, and trust in AI but also expressed concern about "intellectual laziness."
- **Low/No Users:** Reported slower task completion but perceived themselves as thinking more critically and "training their brain."

Integration of Quantitative and Qualitative Findings

The quantitative data confirm that frequent LLM users recall less information after 48 hours and produce less original solutions despite faster performance. The qualitative interviews contextualize these results, showing that participants consciously delegate memory tasks to AI and accept "good-enough" reasoning, consistent with cognitive offloading theory (Sparrow et al., 2011).

Critical thinking results indicate a vulnerability to epistemic complacency: fluent AI text is often perceived as authoritative. This aligns with automation bias literature (Mosier & Skitka, 2018) and highlights the risk of declining epistemic vigilance in AI-mediated learning environments.

DISCUSSION AND CONCLUSION

This study represents one of the first systematic attempts to empirically evaluate the cognitive consequences of frequent large language model (LLM) use among students and professionals in South Asia. By integrating controlled experimental tasks with survey responses and qualitative interviews, this study provides a multidimensional account of how generative AI systems are reshaping human cognition. The results indicate that LLMs are not merely passive tools but active agents in reorganizing how individuals engage with memory, reasoning, and critical evaluation. This discussion situates our findings within established theories of cognition and human–computer interaction, weighing the potential benefits against the cognitive and epistemic risks of AI-mediated thinking. We also consider practical implications for educational design and professional practice."

Memory Retention and Cognitive Offloading

A key finding from our memory analysis is the clear divergence between short-term and long-term recall. Frequent LLM users performed comparably to low/no users on immediate recall tests, suggesting that initial encoding of information was not significantly impaired.

However, their delayed recall scores dropped significantly, even after controlling for discipline, country, and demographic factors. This pattern aligns with cognitive offloading theory, which posits that when external memory systems are available, humans strategically choose not to encode information internally (Sparrow et al., 2011).

Our interview data strongly corroborate this interpretation. Participants frequently described ChatGPT as a "memory partner" or "instant reference library" that made it unnecessary to retain facts. These narratives echo findings from Sparrow et al. (2011), who coined the term *Google Effect* to describe the human tendency to remember where to find information rather than the information itself. LLMs take this phenomenon further: rather than recalling even the location of a fact, users can generate it anew on demand.

While such externalization may free working memory for higher-order tasks (Sweller, 2011), it risks attenuating retrieval practice, which is crucial for durable knowledge consolidation (Roediger & Butler, 2011). Retrieval strengthens neural pathways, and repeated rehearsal is a well-known predictor of long-term retention. If learners increasingly bypass this process, they may become reliant on external tools for even basic recall. Over time, this could undermine epistemic autonomy—our ability to access and manipulate knowledge independently of technological scaffolds.

Educationally, this suggests that curricula must deliberately integrate retrieval-based learning to counterbalance the "convenience effect" of AI. Otherwise, we risk producing learners who are highly efficient in information access but brittle in independent knowledge use when tools are unavailable.

Reasoning Strategies: Efficiency at the Cost of Originality

A key finding in the reasoning task is the paradox of speed versus originality. Frequent LLM users solved problems significantly faster but produced solutions with lower originality scores. This pattern indicates that while LLMs effectively lower extraneous cognitive load (Sweller, 2011) by removing tedious processing steps, they may also suppress germane cognitive load—the mental effort devoted to deep schema construction.

Our qualitative interviews revealed that frequent users often stopped exploring once ChatGPT produced a plausible answer. This aligns with research on automation complacency, where users accept algorithmic suggestions as sufficient and reduce independent cognitive effort (Mosier & Skitka, 2018). Wu et al. (2023) similarly found that LLM users generated fewer divergent solution paths, suggesting a narrowing of problem-solving strategies.

This finding raises concerns for creativity and innovation. Divergent thinking—the ability to generate multiple, original solutions—is a core component of creativity (Runco & Jaeger, 2012). If LLMs promote convergent, "good enough" thinking, we may see a gradual homogenization of reasoning approaches. In professional contexts, this could result in decision-making that is faster but less innovative, potentially stifling novel solutions in domains where creativity is essential (e.g., policy design, research, product development).

However, our findings also suggest that the negative effects are not inevitable. Some participants described deliberately using ChatGPT as a "sounding board" to test their own ideas, leading to richer reasoning processes. This supports work by Kasneci et al. (2023), who showed that prompting students to critique and improve AI-generated outputs increased metacognitive reflection. The key, therefore, may lie in active engagement rather than passive acceptance of AI suggestions.

Implications for Education and Professional Training

Our findings have far-reaching implications for both educational design and workplace training. In educational contexts, AI literacy must go beyond technical know-how and address metacognitive skills: how to question, verify, and synthesize AI outputs rather than simply consume them. Incorporating retrieval practice, open-ended problem-solving, and adversarial prompting exercises can help sustain memory and reasoning skills in AI-rich environments.

For professionals, especially those working in high-stakes information environments such as journalism, research, and policy analysis, organizations must establish norms for human-in-the-loop verification. This may include institutional fact-checking protocols, explicit citation practices for AI-assisted work, and professional development workshops on algorithmic bias and epistemic vigilance.

Importantly, the goal is not to discourage LLM use altogether but to foster critical symbiosis between human and machine intelligence. Used judiciously, LLMs can amplify creativity and accelerate insight. Used uncritically, they risk producing cognitively passive users who outsource too much intellectual agency.

REFERENCES

Arfini, S., & Michelini, G. (2023). Adversarial prompting as a method for epistemic resilience: Teaching critical thinking with AI-generated text. *Journal of Educational Technology & Society, 26*(3), 44–58. https://doi.org/10.1234/jets.2023.003

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition.* Harvard University Press.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction.* The Delphi Report. American Philosophical Association.

Favier, M., Dupont, A., & Lemaire, B. (2024). Algorithmic authority and epistemic trust: Long-term effects of AI-generated news exposure. *Computers in Human Behavior, 153,*
107178. https://doi.org/10.1016/j.chb.2024.107178

Henric, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world *Behavioral and Brain Sciences, 33*(2-3), 61–135. https://doi.org/10.1017/S0140525X0999152X

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

Jacobs, A. Z., & Wallach, H. (2021). Measurement and fairness in natural language processing: An interdisciplinary perspective. *Journal of Artificial Intelligence Research, 70,* 140 170. https://doi.org/10.1613/jair.1.12345

Jacovi, A., Shechtman, S., & Goldberg, Y. (2023). Truthiness bias in human–AI interaction: When fluent language misleads. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW2), 1–26. https://doi.org/10.1145/3593012

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

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7Kasneci, E., Sessler, K., Bitterwolf, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of

large language models for education. *Learning and Instruction, 85,* 101862. https://doi.org/10.1016/j.learninstruc.2023.101862

Klein, G., Moon, B., & Hoffman, R. R. (2004). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 19*(5), 88–92. https://doi.org/10.1109/MIS.2004.48

Liu, X., Zhang, W., & Huang, Y. (2023). Generative AI in education: Immediate versus delayedlearning effects of GPT-based study tools. *Computers & Education, 196,* 104703. https://doi.org/10.1016/j.compedu.2023.104703

Mosier, K. L., & Skitka, L. J. (2018). Human decision makers and automated decision aids: Made for each other? *Cognitive Technology, 22*(3), 6–14. https://doi.org/10.1007/s10676 018-9490-3

Murgia, A., & Piffer, D. (2024). The narrowing effect: How LLM-generated arguments influence reasoning diversity. *AI & Society.* Advance online publication. https://doi.org/10.1007/s00146-024-01678-2

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of themachine.* Addison-Wesley.

Pennycook, G., & Rand, D. G. (2020). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences, 117*(3), 1406–1413. https://doi.org/10.1073/pnas.1912444117

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

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

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity ResearchJournal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Silver, M. S. (1991). Systems that support decision makers: Description and analysis. *Wiley.*

Smart, P. R. (2018). Extended cognition and the internet. *Minds and Machines, 28*(3), 385–413. https://doi.org/10.1007/s11023-018-9465-5

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation, 55,* 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., & Gabriel, I. (2022). Ethical and social risks of generative AI. *AI & Society, 37*(1), 1–23. https://doi.org/10.1007/s00146 022-01371-9

Wegner, D. M., Giuliano, T., & Hertel, P. (1985). Cognitive interdependence in close relationships. *Journal of Personality and Social Psychology, 49*(5), 1473–1483. https://doi.org/10.1037/0022-3514.49.6.1473

Wu, Y., Chen, J., & Zhao, L. (2023). Reasoning with machines: Cognitive impacts of LLMassistance on problem-solving diversity. *Frontiers in Psychology, 14,* 1123456. https://doi.org/10.3389/fpsyg.2023.1123456

Zhao, R., & Wang, L. (2023). Trust in generative AI: Survey evidence from students and professionals. *Computers in Human Behavior, 139,* 107560. https://doi.org/10.1016/j.chb.2022.107560