The Metacognition Paradox

Five Design Principles for AI Tools That Prompt the Thinking They Tend to Replace

TESSA FORSHAW

TFORSHAW@FAS.HARVARD.EDU

HARVARD GRADUATE SCHOOL OF EDUCATION, CAMBRIDGE, MA, USA

Abstract

Metacognition is one of the strongest predictors of learning, performance, and transfer. However, it requires enabling conditions. Conditions such as explicit prompts, dedicated time, occasions that trigger metacognition, and emotional signals influence its occurrence. Generative AI agents, particularly those designed to converge quickly on answers, unintentionally undermine these conditions, creating a Metacognition Paradox. AI erodes the triggers for metacognition at the exact moment when metacognition matters most, because working with AI requires monitoring not just your own thinking but also how you use AI and its outputs. It requires "double metacognition". I identify four enabling conditions that metacognition requires, show how generative AI's default interaction patterns erode each one, and derive five design principles that restore them, reframing metacognition in AI-mediated learning and work as a design problem.

Key Words

Metacognition, Generative AI, Human-AI Interaction

1 INTRODUCTION

Metacognition matters; the evidence is not ambiguous. In Hattie's (2009) synthesis of over 800 meta-analyses, metacognitive strategies rank among the most powerful influences on student achievement. Meta-analyses show that training programs combining cognitive and metacognitive strategy instruction produce the strongest effects on learning (Dignath & Büttner, 2008) and identify metacognition and self-regulation as among the most cost-effective approaches, with an average impact of several additional months of progress (Education Endowment Foundation, 2021). In studies of design education, high-performing students spend significantly more time engaged in metacognitive thought than their lower-performing peers (Kavousi et al., 2020a).

Similarly, the evidence highlights that to gain the most from AI Agent interactions, users need to be metacognitive. In a field experiment with 250 employees using generative AI, metacognitive strategies moderated the extent to which AI use translated into creative gains (Sun et al., 2025). Employees with strong metacognitive strategies saw meaningful increases in cognitive job resources and creativity. Furthermore, AI use improves productivity and work outputs in workers, especially those with strong metacognitive calibration (Caplin et al., 2025).

However, metacognition does not reliably happen on its own. It is more likely to occur when specific conditions support it. First, metacognition must be explicitly prompted. Learners rarely engage in it spontaneously unless it is an established strategy they use (Dignath & Büttner, 2008; Kramarski & Mevarech, 2003). Second, it needs dedicated time and space separate from the cognitive task itself, because metacognition and cognition compete for limited attentional resources (Cuzzolino et al, 2024). Third, people need to recognize that the situation calls for metacognition in the first place, and they often do not, particularly in novel or uncertain contexts where it would be most valuable (Langer, 1989; Perkins et al., 1993). Fourth, emotions function as metacognitive signals; feelings like frustration, confusion, or overconfidence are data about how their thinking is going, but only if they are aware of them (Jia et al., 2019; Kavousi et al., 2020b).

These conditions have always been vulnerable to disruption, particularly when people rely on external tools. For instance, Stone and Storm (2021) found that retrieving information from the internet increased metacognitive bias and reduced metacognitive sensitivity. People became overconfident and less able to distinguish what they knew from what the tool provided. AI Agents amplify this trend, and further undercut the conditions that make metacognition possible in the first place.

Many of AI Agents' default interaction patterns - such as providing comprehensive, polished answers quickly - work directly against the enabling conditions described above. Direct answers remove the occasion for self-prompting. Instant, polished outputs compress the time and space needed for metacognitive pauses. Smoothing away uncertainty and struggle eliminates the cues that signal a need to step back. Moreover, confident, finished products short-circuit the emotional signals, productive frustration, and useful confusion that would otherwise prompt a person to evaluate their approach.

The emerging empirical evidence reflects this. Higher AI tool usage correlates with reduced critical thinking, mediated specifically by increased unintentional cognitive offloading (Gerlich, 2025), and the effects of cognitive offloading on users show up to a 55% reduction in neural connectivity compared to those working without AI assistance (Kosmyna et al., 2025). Perhaps most telling, using ChatGPT on reasoning tasks improved actual performance by 3 points but led to a 4-point overestimation of ability, resulting in a net loss in metacognitive accuracy (Fernandes et al., 2026).

The paradox here is that the design and tendencies of many AI agents are removing the natural conditions and

supports for metacognition at the exact moment they also make metacognition more necessary (Tankelevitch et al., 2024). Yes, because metacognition improves outcomes like creativity and productivity (Sun et al., 2025; Caplin et al., 2025), but also the AI's reasoning as an extension of one's own thinking. Users must now ask themselves questions like: Is the AI's output accurate? Are its assumptions sound? Is it leading me somewhere useful or somewhere convenient? Am I deferring to the AI output as complete when I should be questioning it? How did it get to this output, and how does that compare to my approach?

I call this requirement "double metacognition" - the need to be metacognitive about both one's own thinking and the AI's reasoning under conditions in which AI tools simultaneously erode the supports that typically foster metacognitive regulation. The metacognition paradox is the problem. Double metacognition is what working through it requires, and the design challenge I address here.

2 RELEVANT LITERATURE

2.1 Metacognition and What AI Puts at Risk

To understand what AI endangers, we need to be precise about what metacognition involves in practice. Building on Flavell's (1979) foundational framework, metacognition is understood to involve three core processes: Awareness (noticing what you are thinking and feeling), Evaluation (monitoring whether your current approach is working), and Control (executing strategies to adjust course). Cuzzolino and Grotzer (2022) extend this definition to include awareness, describing, evaluating, and monitoring one's thinking, emotions, motivation, and knowledge, as well as reflecting on and managing the types of thinking tasks required and the social contexts related to one's metacognition. Metacognition extends beyond cognitive processes to include the management of context and the emotional, social, and physical conditions that shape how thinking unfolds (Cuzzolino et al., 2024).

Metacognitive processes do not only happen after the fact. They happen during learning and work, in real time, while people can still influence the outcome (Braden & Forshaw, 2025). This can look like a user noticing mid-brainstorm that she is evaluating ideas too early and deciding to separate generation from critique, or a student realizing he is distracted by jargon and choosing to jot down unfamiliar terms for later so he can keep listening. These are moments when awareness, evaluation, and control operate in the messy middle of the task. It is also exactly what AI removes.

When clinicians use AI-assisted detection over time, they achieve better in-the-moment results, but their unassisted detection rates drop. The active weighing of competing signals that would normally surface doubt gets outsourced, and the skill erodes (Budzyń et al., 2025). When junior developers use AI to write their code, they complete the task but understand less about the concepts they just applied. They got the work done without developing the understanding that would allow them to do it independently next time (Anthropic, 2026).

2.2 Double Metacognition

Often, when we discuss metacognition, we think about it as having a single target: a person's own cognitive processes. However, metacognition has never been limited to monitoring one's own internal cognition. Research on metacognition and tool use shows that evaluating whether an external resource is serving your goals and deciding when to rely on it versus doing the thinking yourself involves the same monitoring, evaluation, and control processes that characterize metacognition about your own thinking (Clarebout et al., 2013; Weis & Wiese, 2019).

The decision to trust or question a tool is a metacognitive judgment directed at something outside your own head. This is consistent with work on social metacognition, which shows that metacognitive judgments operate through the same mechanisms whether directed at one's own cognition or at another agent's (Jost et al., 2012; Shea et al., 2014). With AI, this remains true. There are two targets: the person's own thinking and the AI's, and they interact. They need confidence in their ability to specify what they need, to evaluate whether the AI's output meets that need, and to decide when to automate and when to do the cognitive work themselves (Tankelevitch et al., 2024). Each demands a high level of metacognition.

However, generative AI is not just another tool to be metacognitive about. It combines four features that previous tools did not present together. First, the reasoning behind the outputs is hidden, making them hard to evaluate. Second, the surface quality of those outputs actively signals sound reasoning. Third, unlike a human collaborator, the AI does not convey any genuine signals of uncertainty or doubt. Fourth, unlike previous tools, generative AI simultaneously degrades the very conditions, prompting, time, and emotional cues that would typically support the metacognitive work it demands. The difficulty is compounded by the fact that AI literacy does not appear to help. Higher AI literacy correlates with lower metacognitive accuracy (Fernandes et al., 2026). People who know more about AI are more confident about their performance but less accurate in assessing it.

Double metacognition, then, is not just more metacognition. It requires the ability to hold one's own reasoning alongside the AI's, to notice when one is deferring rather than deciding, and to evaluate outputs one did not produce against standards one may not have fully articulated. It also requires doing this within the very context that undermines it - where the tools people are trying to evaluate are simultaneously reducing the conditions that make evaluation possible. Metacognitive capacities are learnable. However, they will not develop in environments that do not prompt them or in conditions that do not enable them.

Double metacognition is proposed here as a design construct, not solely a cognitive one. The question is not just how individuals can think better with AI, but how AI tools can be designed to sustain the conditions for that thinking.

3 FIVE DESIGN PRINCIPLES FOR DOUBLE METACOGNITION

If we want AI tools that genuinely support learning and thinking, we need to design them differently. The current default, AI that converges quickly on concrete answers, optimizes for output at the cost of cognition. The alternative is AI designed not to complete tasks but to challenge thinking (Sarkar, 2024) and design for metacognition.

Section 1 identified four enabling conditions for metacognition that AI tools tend to undermine. Existing theoretical and empirical research points to how each can be supported in practice. Drawing on that evidence, five principles were developed to restore one or more of these conditions in AI interaction contexts. Several address the same condition through different mechanisms, and they are not claimed to be exhaustive. Together, they form a design brief for tools for thought that protect and prompt double metacognition.

3.1 Elicit Before Generating

If users are to evaluate AI output rather than simply accept it, they need an independent reference point, one that explicit prompting can provide. Before producing its own output, the AI ought to prompt the user for their perspective. Something as direct as "What is your current take on this?" or "What approach are you leaning toward?" prompts the user's reasoning and provides a reference point for evaluating the AI's response. Without this step, the AI's answer becomes the starting point for all subsequent thinking, and metacognition becomes little more than a rationalization of whatever the machine produced.

Research consistently supports the value of prioritizing human thinking. In creative tasks, asking users to share their intentions before engaging with AI leads to richer problem exploration and more effective collaboration (Gmeiner et al., 2025). In learning contexts, requiring learners to explain their understanding before receiving AI feedback yields similar benefits, treating human thinking as the primary object of work (Tomisu et al., 2025). More broadly, when AI asks questions rather than gives answers, people reason more accurately, even when provided with correct explanations (Danry et al., 2023).

An example of this might be when a user asks a generative AI agent to brainstorm solutions to a problem they are working on. Rather than generating a list of ideas, it prompts the user to think and share first: "What ideas have you already considered?

3.2 Show the Reasoning

For users to think critically about AI output, they need access to more than just the final answer; they need to recognize that the situation calls for evaluation. If the AI's reasoning is invisible, users can only evaluate the output itself, which is challenging since a wrong answer can still look polished and therefore play into cognitive biases such as the Dunning-Kruger effect, the Halo effect, the Fluency effect, etc. AI agents should make their assumptions, alternatives considered, and trade-offs explicit. Once they are explicit, users have something to be metacognitive about. They can begin investigating their own thinking by asking questions such as "Do I agree with this assumption?" Would I have framed the problem differently? What did the AI not consider?

Making thinking visible is foundational to metacognitive engagement across problem-solving contexts (Adams et al., 2003; Crismond & Adams, 2012), and when AI reasoning is made explicit, users become more receptive to critically evaluating outputs rather than passively consuming them (Tankelevitch et al., 2024).

In practice, this could look like a user who asks a generative AI agent to recommend a research method. Instead of recommending interviews, it responds: "I would suggest interviews over surveys here because your questions are exploratory, but surveys would give you broader reach if that matters more. What is the priority?"

3.3 Pause in the Messy Middle

The process of working through a problem, getting stuck, realizing an approach is not working, and reframing is where the time, space, and emotional signals for metacognition naturally arise. AI agents that skip straight to a final output remove that generative struggle entirely. By eliminating the messy middle, they remove the cognitive handholds that allow people to monitor, evaluate, and modify their thinking as it develops. Gmeiner et al. (2025) describe this as reduced cognitive involvement, in which users experience insufficient exploration of the problem when AI produces finished outputs.

However, preserving the messy middle is not enough on its own. Metacognition is cognitively expensive and cannot run continuously alongside the primary task without degrading both the primary task and metacognition (Cuzzolino et al., 2022). People need deliberate pauses within the process, not just the process itself. The most productive moments are decision points, junctures where the direction of thinking could meaningfully change. AI agents should embed targeted prompts at critical moments, such as before committing to an approach, after receiving unexpected results, and/or when a clear moment of confusion arises. The dosage of these instances also matters. When metacognitive scaffolding is consistent with learners' immediate goals, it increases the use of self-regulated learning strategies. The goal is strategic pauses that create space for monitoring and evaluation, rather than continuous scaffolding that risks replacing the very capacity it aims to develop.

In application, this might show up when a user asks an AI agent to develop a class project plan. Instead of delivering a finished plan, it shares an early rough outline with gaps and open questions marked, and asks the user: "Here is where my thinking is so far. What is missing?"

3.4 Surface Uncertainty Before Converging

AI outputs carry false confidence at two levels. At the claim level, everything sounds equally authoritative regardless of how well-supported it is, and without confidence calibration,

users' confidence ratings become disconnected from actual accuracy (Fernandes et al., 2026). At the output level, a polished AI output often sends users a misleading emotional signal that the work is done. This matters because one of the most compelling findings in the metacognition literature is that people fail to notice the occasions that call for deliberate, reflective thinking (Perkins et al., 1993), and AI interaction compounds this problem, since a confident tone can mask genuine uncertainty. Because users cannot be metacognitive about uncertainty they do not know exists, false confidence leads directly to premature closure. A person under time pressure or cognitive load may accept an output that merely looks adequate (Kavousi et al., 2020b; Butler & Roberto, 2018) and have an inflated sense of confidence in their own understanding (Stone & Storm, 2021).

AI agents ought to flag moments when confidence is low, when the problem is ambiguous, and when assumptions could go either way, rather than presenting every output with equal authority. Equally, they should build friction at moments of apparent completion, prompting users to compare what they asked for against what they received, to check whether the output addresses the actual problem or a convenient simplification of it.

In application, this might show up when synthesizing field research: a generative AI agent flags where participant accounts conflict or where the data is thin, rather than presenting a confident narrative. Or when a user asks to proceed with a concept it has drafted, it prompts: "Does this really solve the problem as you framed it at the start? Why or why not?"

3.5 Make Help-Seeking Metacognitive

When users turn to AI for help, the interaction is almost always transactional. Help that solves the problem for the learner produces different outcomes than help that guides the learner to solve it themselves (Nelson-Le Gall, 1985), and AI overwhelmingly defaults to solving the problem. However, help-seeking is itself a metacognitive strategy, one that requires noticing the limits of your own understanding, identifying what specifically you need, and evaluating whether the help you receive actually addresses the gap (Dignath & Büttner, 2008).

Since help-seeking is a metacognitive strategy, deciding when and how to use a tool to aid thinking is itself a metacognitive skill (Clarebout et al., 2013). The decision to turn to AI in a given moment requires the user to assess whether they have reached the limits of their own thinking, whether they are offloading prematurely, or if there is value in offloading the task. AI agents should make use of these moments. Before answering, the AI could prompt: "What do you already know about this? Where exactly are you stuck? What kind of help are you seeking and why?" When AI is designed to prompt reasoning and reflection rather than provide direct answers, it reframes the interaction from passive consumption to active self-assessment, turning a moment of cognitive outsourcing into a moment of metacognitive engagement (Mollick & Mollick, 2023).

4 DISCUSSION AND FUTURE DIRECTIONS

I reframe metacognition in AI-mediated learning and work as a design problem. Prior work has documented the metacognitive demands generative AI creates and rightly called for AI tools that support rather than undermine thinking. I continue in that vein, and contribute a specific mechanism for doing so: identifying the enabling conditions metacognition requires, showing how current AI tools erode them, and deriving design principles that restore them.

Introducing double metacognition shows that generative AI extends metacognitive demands beyond a person's own thinking to include the AI's reasoning, while simultaneously undermining the conditions that typically support metacognitive regulation. Because metacognition is a design problem, the responsibility for solving it extends beyond individual users. The principles proposed here have implications for tool builders shaping default interaction patterns, for educators teaching with AI tools, and for organizations deploying them at scale. Design choices that erode metacognition at the individual level become systemic risks.

Each of the five design principles targets a specific enabling condition that many current AI tools tend to remove. However, it would be remiss not to highlight that they are framed around convergent AI tools such as chatbots, code generators, and writing assistants. More exploratory AI tools designed for brainstorming or open-ended ideation may interact with metacognition differently, and the principles may need to be adapted in those contexts.

There is also a tension at the heart of the proposal. These principles work by adding friction. However, friction that becomes predictable risks becoming performative. If metacognitive prompts are experienced as obstacles rather than genuine invitations to think, users will dismiss them. Designing for metacognition is not just about whether prompts are present but about whether the conditions make genuine engagement likely.

Proposing principles is also only a first step. The next step is to test whether they work and to uncover which specific embodiments of the principles, if any, are effective. Prior research tells us that AI can create a gap between how well we think we are doing and how well we actually perform. Can design principles like Elicit Before Generating help close that gap? Does pausing in the messy middle at decision points help people build lasting habits that mitigate the risk?

Effective use of generative AI that limits its impact on cognition and performance overwhelmingly demands more metacognition from users at the very moment it removes the conditions that support it.

ACKNOWLEDGMENTS

Thank you to colleagues at the Next Level Lab at Project Zero and the Harvard Graduate School of Education for years of discussion about metacognition, and Rich Braden for exploring these ideas together in the classroom. Claude was used for editing and feedback on drafts, and Grammarly for proofreading.

REFERENCES

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. Design Studies, 24(3), 275–294. https://doi.org/10.1016/S0142-694X(02)00056-X

Braden, R., & Forshaw, T. (2025). Innovation-ish: How anyone can create breakthrough solutions to real problems in the real world. Wiley

Budzyń, K., Bisschops, R., & Kamiński, M. F. (2025). Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. The Lancet Gastroenterology & Hepatology, 10(10), 896–903. https://doi.org/10.1016/S2468-1253(25)00133-5

Butler, A. G., & Roberto, M. A. (2018). When cognition interferes with innovation: Overcoming cognitive obstacles to design thinking. Research-Technology Management, 61(4), 45–51. https://doi.org/10.1080/08956308.2018.1471276

Clarebout, G., Elen, J., Collazo, N.A.J., Lust, G., & Jiang, L. (2013). Metacognition and the use of tools. In R. Azevedo & V. Aleven (Eds.), International handbook of metacognition and learning technologies (pp. 187–195). Springer. https://doi.org/10.1007/978-1-4419-5546-3_13

Caplin, A., Deming, D., Li, S., Martin, D., Marx, P., Weidmann, B., & Ye, K. J. (2025). The ABCs of who benefits from working with AI: Ability, beliefs, and calibration. Management Science. Advance online publication. https://doi.org/10.1287/mnsc.2024.08994

Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. Journal of Engineering Education, 101(4), 738–797. https://doi.org/10.1002/j.2168-9830.2012.tb01127.x

Cuzzolino, M. P., & Grotzer, T. A. (2022). The icing on the cake: How metacognition enhances learning. Next Level Lab, Harvard Graduate School of Education.

Cuzzolino, M. P., Sun, M., Xu, J., Becerra, J., Fields, E., & Grotzer, T. A. (2024). Leveraging the power of metacognition and contextualized agency for workplace learning. Paper presented at NEERO 2024.

Danry, V., Pataranutaporn, P., Mao, Y., & Maes, P. (2023). Do not just tell me, ask me: AI systems that intelligently frame explanations as questions improve human logical discernment accuracy compared to causal AI explanations. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). ACM. https://doi.org/10.1145/3544548.3580672

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. Metacognition and Learning, 3(3), 231–264. https://doi.org/10.1007/s11409-008-9029-x

Education Endowment Foundation. (2021). Metacognition and self-regulation. Teaching and Learning Toolkit. https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/metacognition-and-self-regulation

Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2026). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. Computers in Human Behavior, 175, 108779. https://doi.org/10.1016/j.chb.2025.108779

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

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

Gmeiner, F., Luo, K., Wang, Y., Holstein, K., & Martelaro, N. (2025). Exploring the potential of metacognitive support agents for human-AI co-creation. In Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25) (pp. 1244–1269). Association for Computing Machinery. https://doi.org/10.1145/3715336.3735785

Hargrove, R. A., & Nietfeld, J. L. (2015). The impact of metacognitive instruction on creative problem solving. The Journal of Experimental Education, 83(3), 291–318. https://doi.org/10.1080/00220973.2013.876604

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge.

Jia, X., Li, W., & Cao, L. (2019). The role of metacognitive components in creative thinking. Frontiers in Psychology, 10, 2404. https://doi.org/10.3389/fpsyg.2019.02404

Jost, J.T., Kruglanski, A.W., & Nelson, T.O. (2012). Social metacognition: An expansionist review. In P. Brinol & K.G. DeMarree (Eds.), Social metacognition (pp. 17-41). Psychology Press. https://doi.org/10.4324/9781315799278

Karabenick, S. A., & Gonida, E. N. (2017). Academic help-seeking as a self-regulated learning strategy. In D. H. Schunk & J. A. Greene (Eds.), Handbook of self-regulation of learning and performance (2nd ed., pp. 421–433). Routledge.

Kavousi, S., Miller, P. A., & Alexander, P. A. (2020a). Modeling metacognition in design thinking and design making. International Journal of Technology and Design Education, 30(4), 709–735. https://doi.org/10.1007/s10798-019-09521-9

Kavousi, S., Miller, P. A., & Alexander, P. A. (2020b). The role of metacognition in the first-year design engineering lab. Educational Technology Research and Development, 68(6), 3471–3494. https://doi.org/10.1007/s11423-020-09848-4

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

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. American Educational Research Journal, 40(1), 281–310. https://doi.org/10.3102/00028312040001281

Nelson-Le Gall, S. (1985). Help-seeking behavior in learning. Review of Research in Education, 12(1), 55–90. https://doi.org/10.3102/0091732X012001055

Mollick, E., & Mollick, L. (2023). Assigning AI: Seven approaches for students with prompts (SSRN Working Paper No. 4475995). Wharton School, University of Pennsylvania. https://doi.org/10.2139/ssrn.4475995

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. Merrill-Palmer Quarterly, 39(1), 1–21.

Sarkar, A. (2024). AI should challenge, not obey. Communications of the ACM, 67(10), 18–21. https://doi.org/10.1145/3649404

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C.D. (2014). Supra-personal cognitive control and metacognition. Trends in Cognitive Sciences, 18(4), 186–193. https://doi.org/10.1016/j.tics.2014.01.006

Stone, C.B., & Storm, B.C. (2021). Distributed metacognition: Increased bias and deficits in metacognitive sensitivity when retrieving information from the internet. Technology, Mind, and Behavior, 2(3). https://doi.org/10.1037/tmb0000040

Shen, J. H., & Tamkin, A. (2026). How AI impacts skill formation. arXiv preprint. https://arxiv.org/abs/2601.20245

Sun, S., Li, Z. A., Foo, M.-D., Zhou, J., & Lu, J. G. (2025). How and for whom using generative AI affects creativity: A field experiment. Journal of Applied Psychology, 110(12), 1561–1573. https://doi.org/10.1037/apl0001296

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., & Sellen, A. J. (2024). The metacognitive demands and opportunities of generative AI. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). ACM. https://doi.org/10.1145/3613904.3642902

Tomisu, H., Ueda, J., & Yamanaka, T. (2025). The cognitive mirror: A framework for AI-powered metacognition and self-regulated learning. Frontiers in Education, 10, 1697554. https://doi.org/10.3389/feduc.2025.1697554

Weis, P.P., & Wiese, E. (2019). Using tools to help us think: Actual but also believed reliability modulates cognitive offloading. Human Factors, 61(2), 243–254. https://doi.org/10.1177/0018720818797553