Cognitive Hygiene: The Anatomy of AI-Assisted Thinking

A narrative literature review on LLMs and independent, critical, collaborative, and creative thought

Prepared: July 2026

Scope: Research on large language models, generative AI, cognitive offloading, automation bias, AI-assisted learning, human-AI collaboration, and creativity support.

Review model: This review follows the broad strategy of Peter Kollock's narrative review *Social Dilemmas: The Anatomy of Cooperation*: it begins with a central problem, maps the recurring models and metaphors, then turns to solution families and future directions. Source document.

Key words: cognitive hygiene, large language models, generative AI, cognitive offloading, automation bias, critical thinking, independent thought, collaboration, creativity, learning, human-AI interaction.

Abstract

The study of cognitive hygiene is the study of the tension between cognitive efficiency and cognitive agency. LLMs make it individually rational to ask a machine to summarize, draft, explain, brainstorm, debug, role-play, or decide. Often this works: experiments in writing, consulting, customer support, tutoring, and ideation show gains in speed, average quality, access to expertise, and sometimes learning. But the same convenience can also produce a deficient equilibrium. If people repeatedly skip the effort that builds memory, judgment, skepticism, taste, and shared understanding, they may become more productive at the moment of use while less capable when the tool is absent or wrong.

The evidence is mixed because LLMs are not a single intervention. They are a family of cognitive environments whose effects depend on task, expertise, stakes, interaction design, and time horizon. Unguarded answer-giving can become a crutch; scaffolded tutoring can improve learning. AI suggestions can lift weak writers and generate useful ideas; shared model defaults can also homogenize style, culture, and collective novelty. AI can provoke better critique; it can also reduce the perceived need for critical effort, make false explanations persuasive, and encourage overreliance. Cognitive hygiene, then, is not a policy of abstinence. It is the practice of preserving human authorship over questions, standards, evidence, and final judgment while using LLMs as calculators, tutors, collaborators, and muses only where those roles are earned.

Contents

All author-year citations in the review and all entries in the source list are linked to source documents, publisher pages, DOIs, preprints, or official project pages where available.

1. The Question of Cognitive Hygiene

Cognitive hygiene asks how people can use systems that produce fluent language without surrendering the habits that make thought trustworthy: forming one's own questions, testing claims, remembering what matters, noticing uncertainty, listening to others, and generating ideas that are not merely average.

Large language models have made a familiar bargain much more tempting. We have long used tools to offload cognition: notebooks, calculators, search engines, spreadsheets, checklists, maps, templates, and collaborators. The literature on cognitive offloading describes this as using physical or external action to reduce internal processing demands. Offloading is not inherently bad. It is part of civilization. A clean calendar, a reliable checklist, and a well-designed search process can extend memory and attention. The hygiene problem begins when the external system does not merely store or calculate, but supplies plausible language that can stand in for understanding.

LLMs sit at an unusual point in the history of cognitive tools. They are not only repositories or calculators; they are conversational producers of explanations, summaries, plans, arguments, code, feedback, stories, and social language. Their outputs often feel like the work of a knowledgeable partner. They can also be wrong, generic, overconfident, culturally flattening, or optimized for fluency rather than truth. The result is a new class of cognitive dilemma. A person who uses an LLM for every first draft may produce more pages today while getting less practice at planning, phrasing, and revising. A student who receives immediate solutions may finish homework while learning less. A team that uses one model as an ideation partner may converge quickly on acceptable ideas while losing diversity of approaches. These are not simple stories of decline; they are tradeoffs between performance at the moment of use and capability after use.

This review is selective. It emphasizes empirical work and adjacent theory most relevant to four abilities named in the prompt: independent, critical, collaborative, and creative thinking. The term *cognitive hygiene* is treated as an umbrella rather than a settled research field. The studies come from human-computer interaction, cognitive psychology, education, management, economics, learning science, creativity research, and automation. The aim, like Kollock's review of social dilemmas, is not to list everything but to provide a structure for understanding the area and a set of pointers to useful resources.

> **A recurring distinction:** the central question is not simply whether an LLM improves output. It is whether the process of using the LLM leaves the human, pair, team, or institution more capable of thinking well in the next situation.

2. Necessary, Dangerous Metaphors

Kollock warned that the metaphors that organize a literature can be both necessary and dangerous. They help a field think, but they can also blind it. LLMs have already collected a set of metaphors - assistant, calculator, tutor, colleague, oracle, muse, prosthesis, intern, autocomplete, and second brain. Each captures one part of the phenomenon and distorts another.

The calculator metaphor

When the task is repetitive, well-specified, and easily checked, the calculator metaphor is powerful. A model can convert formats, produce boilerplate, outline possibilities, or summarize a routine thread, freeing attention for higher-order work. This is one way to understand productivity studies such as Noy and Zhang (2023), who found that ChatGPT substantially reduced time on professional writing tasks while improving average output quality, and Brynjolfsson, Li, and Raymond (2025), who found productivity gains for customer support agents using a generative AI assistant.

But the calculator metaphor becomes dangerous when the task is not merely computation but judgment formation. People can learn arithmetic with a calculator, but not if the calculator always replaces the operations the learner is meant to practice. In the same way, a model can support writing, reasoning, or coding while still interrupting the practice loops that build skill.

The tutor metaphor

The tutor metaphor is also both promising and dangerous. Carefully designed AI tutoring can create adaptive explanations, generate examples, and keep students engaged. Kestin et al. (2025) report that an AI tutoring approach can outperform in-class active learning on immediate learning measures when the interaction is designed around pedagogical best practices. In contrast, Bastani et al. (2025) found that generative AI without guardrails can harm learning: students using unguarded GPT-4 performed worse later when assistance was removed, while a tutor-style, scaffolded version mitigated the harm.

The contrast is instructive. A tutor asks, waits, diagnoses, and provides help at the edge of the learner's current ability. An answer machine removes the edge. Cognitive hygiene requires preserving that edge.

The colleague metaphor

LLMs can also act like colleagues who contribute drafts, critique plans, or suggest alternatives. In organizations, this can spread tacit best practices. Brynjolfsson and colleagues interpret their customer support findings partly as a mechanism by which AI helps lower-skill workers access the behaviors of more experienced workers. Dell'Acqua et al. (2026) similarly show strong performance gains for consultants on tasks inside the model's capability frontier, but losses when people use the model for

tasks outside that frontier. A colleague metaphor is hygienic only if the human still understands what the colleague is good at, what the colleague cannot see, and who remains accountable.

The oracle metaphor

The oracle metaphor is the most seductive and the most dangerous. LLMs are fluent, immediate, and often correct enough to invite deference. Yet pre-LLM automation research already showed that people can over-trust automated aids. Parasuraman and Manzey (2010) reviewed automation complacency and bias; recent AI-advice experiments such as Klingbeil, Grutzner, and Schreck (2024) extend that concern to contemporary AI contexts. The oracle metaphor erodes hygiene when fluency substitutes for evidence.

The muse metaphor

Finally, the muse metaphor captures a real creative benefit. A model can offer prompts, variations, analogies, and unexpected combinations. Doshi and Hauser (2024) found that generative AI can improve the creativity of individual short stories, especially for less creative writers. But the muse is shared. If many people prompt similar models in similar ways, the collective result may be more polished and less varied. The same study found that AI-assisted stories became more similar to one another. The muse becomes a monoculture when everyone hears the same whisper.

---

3. Modeling LLM-Mediated Cognition

The effects of LLM use are not well modeled by a single yes-or-no question. The literature points toward a small set of dimensions that recur across domains.

The task frontier

One of the most useful concepts is the "jagged technological frontier" in Dell'Acqua et al.. The idea is simple but powerful: AI is not uniformly good or bad at professional work. It can be excellent at some tasks that look difficult and unreliable at some tasks that look similar. For cognitive hygiene, this means that users must develop frontier literacy - the ability to know when a model is likely to help, when it is likely to mislead, and when human judgment must dominate.

User expertise and self-confidence

Expertise changes the meaning of assistance. A novice may benefit from examples and structure but may also lack the knowledge to detect errors. An expert may use the same output as raw material, not authority. This pattern appears in studies of critical thinking and AI reliance. Lee et al. (2025) found that workers' confidence in AI was

associated with less self-reported critical thinking effort, while confidence in their own ability was associated with more effort. The danger is not reliance alone; it is reliance without enough self-efficacy or domain knowledge to challenge the system.

Interaction mode

LLM use can be answer-receiving, question-answering, critique-seeking, co-drafting, simulating, tutoring, or role-playing. These modes are cognitively different. A model that gives an answer can reduce effort. A model that asks questions can increase effort. Work on question-framed explanations by Danry et al. (2023) and provocation-based AI assistance by Drosos et al. (2025) suggests that interaction design can move users from passive acceptance toward metacognition and critique.

Time horizon

Some studies measure performance during AI use. Others measure what people can do after assistance is removed. This distinction often explains apparent contradictions. Immediate performance can improve while later unaided performance declines. Bastani et al.'s learning study is important because it observes this after-use dimension directly. Cognitive hygiene is primarily about the longer horizon.

Level of analysis

An intervention can help an individual but harm a group, or help a group but reduce cultural diversity. The creativity literature makes this especially clear. Individual stories can improve while the set of stories becomes more similar. A team can converge faster while the broader organization loses dissenting frames. This is where the analogy to social dilemmas is strongest: individually reasonable LLM use can create collective sameness.

Dimension	Hygienic use	Unhygienic drift
Task frontier	Use AI where outputs are checkable or within known strengths.	Use fluent output as if all tasks are equally safe.
Expertise	Use AI to extend skills and compare against one's own answer.	Use AI to replace the practice that builds skills.
Interaction mode	Ask for questions, critiques, alternatives, evidence, and uncertainty.	Ask for final answers, then lightly edit.
Time horizon	Evaluate both output quality and later unaided capability.	Measure only speed or polish at the moment of assistance.
Level of analysis	Protect diversity of drafts, sources, and voices.	Let a shared model converge everyone toward the same defaults.

Table 1. A cognitive hygiene reading of recurring dimensions in the LLM literature.

4. Independent Thinking

Independent thinking does not mean thinking alone. Human thought has always depended on tools, teachers, books, and other people. Independence means preserving authorship over the question, the standard of adequacy, and the final judgment. The LLM literature shows both ways this can be strengthened and ways it can be weakened.

Positive impacts: access, rehearsal, and private scaffolding

For many users, the first benefit of LLMs is lowered friction. A blank page becomes an outline; a vague question becomes a set of candidate formulations; a confusing concept becomes a sequence of examples. In Noy and Zhang's writing experiment, generative AI made professional writing faster and improved measured quality. In workplaces, AI can provide lower-performing workers with examples of what more experienced workers might do, as seen in Brynjolfsson et al.'s customer support study. In learning contexts, scaffolded AI tutors can create more opportunities for practice and feedback than many classrooms can provide.

There is a democratic promise here. LLMs can give a person access to a tireless interlocutor, editor, or explainer. They can make it easier to ask naive questions privately. They can help users rehearse arguments, translate between registers, or see a domain from multiple perspectives. For people who lack expert networks, time, confidence, or language fluency, this can support rather than undermine independence. A person can become more independent when a tool helps them enter a conversation from which they were previously excluded.

Negative impacts: substitution, deference, and voice capture

The danger is that assistance becomes substitution. Bastani et al.'s study gives this concern an unusually clear form: unguarded access to GPT-4 helped students complete practice but later hurt unaided performance. This is the crutch problem. The student appears more capable while supported, but the support has displaced the cognitive work that produces durable capability.

Independence can also be weakened by deference. Automation-bias research shows that people may accept automated recommendations even when they have reasons to doubt them. Klingbeil and colleagues find costly overreliance on AI advice even when it conflicts with contextual information. Lee et al. add that confidence in AI is associated with reduced critical thinking effort. These findings do not mean that users always obey models. They mean that model fluency changes the cost and social feel of dissent. It is easier to accept a polished suggestion than to defend an awkward doubt.

A subtler threat is voice capture. Jakesch et al. (2023) found that opinionated language-model co-writing could influence both what participants wrote and their later attitudes. The finding matters because independence is not only about factual correctness. It is also about the formation of preference, stance, and language. A model that supplies the words may also supply the frame.

> **Independent thinking, hygienically defined:** the user can say what they thought before seeing the model, what changed after seeing it, why the change was warranted, and what evidence would make them change back.

---

5. Critical Thinking

Critical thinking is the family of practices by which people examine claims, compare evidence, identify assumptions, notice alternatives, and calibrate confidence. LLMs can support these practices because they can generate counterarguments and explanations on demand. They can also weaken them because they produce plausible answers quickly enough to make further scrutiny feel unnecessary.

Positive impacts: criticism on demand

At their best, LLMs provide cheap access to adversarial reasoning. A user can ask for counterexamples, alternative hypotheses, hidden assumptions, stakeholder objections, statistical caveats, or likely failure modes. This can help users who would otherwise stop at their first coherent answer. It can also help teams separate idea generation from evaluation by asking the model to play a critic after humans have produced an initial view.

Several HCI studies point toward design patterns that preserve critical effort. Danry et al. found that framing AI explanations as questions can improve logical discernment compared with simply presenting causal explanations. Drosos et al. tested AI-generated provocations in knowledge work and found that textual critiques and prompts can elicit more critical and metacognitive engagement. These studies are important because they shift the model from answer-giver to thinking partner. The model is not trusted less by being made useless; it is made useful in a way that keeps the human active.

Negative impacts: reduced effort, overgeneralization, and persuasive reasons

The same fluency that makes a model useful makes it risky. Lee et al.'s survey of knowledge workers found that users reported applying critical thinking in many AI-assisted tasks, but also that higher confidence in AI related to lower perceived critical thinking effort. This suggests a hygiene gradient: as the system feels more competent, users may shift from scrutiny to supervision, and from supervision to acceptance.

One specific risk is overgeneralized synthesis. Peters and Chin-Yee (2025) studied LLM summaries of scientific research and found substantial overgeneralization compared with human-written summaries. This matters for literature reviews, policy memos, and scientific communication, where the difference between "this study found" and "research shows" is often the difference between accurate synthesis and misleading authority.

Another risk is that explanations can make false claims more persuasive. Danry et al. (2024) experimentally examined deceptive AI systems that provide explanations and found that explanations can amplify the persuasiveness of false outputs. This fits a broader psychological concern: reasons can create a feeling of transparency even when the underlying system is opaque or wrong. In LLM contexts, a bad answer plus a fluent rationale can be more dangerous than a bad answer alone.

Critical thinking with LLMs therefore requires a different posture from ordinary reading. The user must examine not only the claim, but the relationship between fluency, evidence, and uncertainty. A well-formed paragraph is not a warrant. A list of reasons is not a proof. A citation is not evidence until it is checked.

---

6. Collaborative Thinking

Collaborative thinking is not merely the sum of individual thoughts. It includes coordination, mutual correction, role differentiation, common ground, conflict, trust, and shared standards. LLMs can improve collaboration by helping people externalize ideas and access expertise. They can also degrade collaboration by inserting a persuasive, unaccountable participant into the group.

Positive impacts: coordination and access to tacit practice

In workplaces, LLMs can help groups coordinate around shared drafts, summaries, decision logs, and alternative framings. Brynjolfsson et al.'s customer support study suggests one mechanism: AI can capture and redistribute practices associated with more experienced or more effective workers. In consulting, Dell'Acqua et al. found large gains on tasks inside the technological frontier, indicating that AI can act as a powerful production partner when task demands match model strengths.

Educational collaboration studies also report benefits. Wei et al. (2025) found that generative AI support improved collaborative problem-solving and team creativity in a digital story creation setting. Shaer et al. (2024) studied AI-augmented brainwriting and explored how LLMs can support group ideation and evaluation. Earlier co-writing infrastructure such as CoAuthor also showed that people use language models not only to finish sentences, but to explore language, ask for ideational help, and manage the rhythm of joint writing.

These benefits matter because many groups fail not from lack of intelligence but from lack of process. LLMs can help turn tacit reasoning into visible text. They can make meetings more accessible to non-native speakers, quieter participants, or people who need time to formulate ideas. They can provide a common artifact around which disagreement can organize.

Negative impacts: false consensus, accountability gaps, and displaced human interaction

The same shared artifact can also become a false center. If a group treats the model's output as the neutral starting point, the model sets the frame. The cost of objecting rises because the objection appears to be against efficiency, not merely against a claim. In this way, AI can become an unaccountable third member of the group: influential, tireless, and never responsible.

Overreliance has social costs. Klingbeil et al. emphasize that poor reliance on AI advice can harm third parties, not only the person using the tool. This is especially relevant in organizational settings where AI-assisted decisions affect customers, students, patients, workers, or applicants. A team may believe it is exercising oversight while in practice rubber-stamping a system whose errors no one owns.

There is also a risk of displaced human interaction. If a student asks the model instead of a peer, a writer asks the model instead of an editor, or a worker asks the model instead of a colleague, they may lose the friction through which communities build shared standards. Human disagreement carries social information. It reveals values, stakes, histories, and forms of expertise that the model may not represent. Collaborative hygiene therefore requires deliberate moments when humans think together before the model summarizes, smooths, or reconciles.

---

7. Creative Thinking

Creativity research is where the double effect of LLMs appears most vividly. LLMs can help individuals produce more ideas and more polished artifacts. At the same time, they can reduce collective diversity by pulling many users toward similar patterns. The result is an apparent paradox: more creativity for the individual, less novelty for the group.

Positive impacts: ideation, variation, and quality lifting

Generative AI is well suited to divergent prompting. It can produce many variations, analogies, constraints, audiences, titles, product concepts, story openings, or visual directions. In innovation research, Girotra, Meincke, Terwiesch, and Ulrich compared LLM-generated ideas with human ideas and found that LLMs could produce high-quality candidate ideas at scale. Boussioux et al. (2024) found that human-AI combinations performed well on several dimensions of creative problem solving, even while human-only crowds retained advantages on novelty.

Doshi and Hauser's short-story experiment is especially important because it separates individual and collective levels. Access to generative AI improved the rated creativity and quality of stories, with larger benefits for less creative writers. This is a genuine positive impact. LLMs can raise the floor. They can help users escape the first idea, find a structure, and add texture.

Negative impacts: homogenization, cultural flattening, and taste atrophy

The danger is that raising the floor can lower the variance. Doshi and Hauser found that AI-assisted stories were more similar to each other. Moon et al. (2025) report a similar homogenizing pattern in a large comparison of human and ChatGPT-assisted admissions-style writing. Agarwal, Naaman, and Vashistha (2025) add a cultural dimension: AI writing suggestions can pull users toward more Westernized styles and reduce cultural nuance.

This is not simply an aesthetic worry. Creativity depends on the ecology of differences: different memories, metaphors, dialects, disciplinary habits, jokes, mistakes, constraints, and tastes. If model suggestions become the common substrate of many creative acts, collective novelty may decline even as average acceptability rises. The risk is not that AI makes bad art. The risk is that it makes many artifacts good in the same way.

There is also a developmental issue. Creative skill is not only the ability to generate options; it is the ability to form taste through struggle. A model can offer ten alternatives instantly, but the user's taste develops by noticing why one alternative is alive and another is merely plausible. Cognitive hygiene in creative work means using the model to widen the field while preserving the human labor of selection, revision, and owning a style.

---

8. Practicing Cognitive Hygiene

Kollock grouped solutions to social dilemmas into motivational, strategic, and structural families. A similar organization is useful here. Motivational solutions change what users value and attend to. Strategic solutions change how users interact with the model. Structural solutions change tools, norms, assessments, and institutions so that good cognitive practice is easier than bad practice.

Motivational solutions: preserve authorship and self-efficacy

The first family of solutions begins before the prompt. Users need a norm of authorship: state your question, hunch, criteria, or confusion before asking the model. This can be as simple as writing a one-paragraph "my current view" note before requesting feedback. The goal is not to be right first. It is to create a trace of independent thought so that later changes can be evaluated rather than absorbed.

Motivational hygiene also requires self-efficacy. Lee et al.'s findings suggest that people who feel more confident in their own ability engage in more critical thinking around AI. This does not mean inflating confidence. It means preserving the experience of competence through practice. Students, writers, analysts, and teams need recurring AI-free attempts, not because AI is forbidden, but because unaided attempts reveal what help is needed and keep the person from mistaking the model's fluency for their own mastery.

Strategic solutions: use the model as critic, not only producer

The second family concerns interaction strategy. The hygienic shift is from "produce the answer" to "improve my thinking." Useful prompts ask the model to identify assumptions, generate counterexamples, ask diagnostic questions, supply competing interpretations, or critique a human draft against explicit criteria. This is the practical lesson of question-framed explanations and provocation-based systems: keep the human in the loop not as a passive approver, but as an active reasoner.

Several routines follow from this principle. First, attempt the task before asking for the model's solution when the goal is learning or judgment development. Second, require evidence and source checking when the task depends on facts. Third, separate divergent and convergent phases: brainstorm without AI, then use AI to expand; or brainstorm with AI, then evaluate without AI. Fourth, ask for uncertainty and failure modes, but do not treat the model's self-critique as verification. Fifth, compare multiple sources or models when collective diversity matters.

> **A simple hygiene protocol:** Try first. Ask for critique. Verify externally. Revise in your own words. Record what changed and why.

Structural solutions: design tools and institutions around thinking, not only output

Individual willpower is not enough. The most important hygiene interventions are structural. Tool designers can default to questions, uncertainty, provenance, and calibrated confidence rather than finished prose. Systems can make it easy to see source documents, compare claims, preserve drafts, and flag when users are making high-stakes decisions from model-generated summaries. They can also introduce productive friction: "What is your current answer?" before revealing a solution, or "Which source supports this claim?" before exporting a report.

Educators can assess process as well as product. A course that permits LLM use can require students to submit initial attempts, prompt logs, source checks, and reflections on what assistance changed. Workplaces can do the same for decisions: record the human rationale, the AI contribution, the verification step, and the accountable owner. In creative organizations, teams can preserve human-only ideation rounds before model expansion, rotate models or prompts to avoid monoculture, and evaluate diversity rather than only average quality.

Finally, institutions should measure after-use capability. If a tool improves assisted output but lowers later unaided performance, that is not a simple productivity gain. The relevant metric is not only how much work was completed, but what capacities were built, maintained, or lost.

---

9. Conclusions and Future Directions

The LLM literature does not support a simple verdict. It supports a conditional one. LLMs can be powerful cognitive scaffolds. They can also become cognitive shortcuts that bypass the very work they appear to enhance. Positive effects are strongest when tasks are within the model's frontier, when outputs are checkable, when users have enough expertise to evaluate suggestions, and when the model is designed to ask, provoke, and scaffold rather than merely answer. Negative effects are most concerning when users are novices, tasks are hard to verify, the model is treated as an oracle, or many people use the same model defaults in domains where diversity matters.

Several research gaps follow. First, the field needs more longitudinal studies that measure skill after assistance is removed. Second, it needs better measures of cognitive hygiene itself: independent problem formulation, calibration, error detection, transfer, retention, originality, and diversity. Third, studies should compare interaction modes, not just model access versus no access. The difference between an answer machine and a Socratic tutor may be larger than the difference between two model versions. Fourth, the field needs collective-level measures. Average individual quality is not enough when the outcome of concern is cultural or intellectual homogenization.

The near-term conclusion is practical. Use LLMs, but do not let them occupy every cognitive role at once. Use them as calculators for checkable transformations, tutors that ask before telling, colleagues whose work is reviewed, critics that widen scrutiny, and muses that expand possibilities. Do not use them as oracles of final truth, substitutes for practice, or universal first authors. The core hygiene rule is simple: preserve the friction where learning, judgment, collaboration, and originality are formed.

Linked Source Documents

The list below emphasizes sources discussed in the review. Publisher pages, DOI links, preprints, or official project pages are provided where available.

Agarwal, D., Naaman, M., & Vashistha, A. (2025). *AI suggestions homogenize writing toward Western styles and diminish cultural nuances*. CHI 2025. Source document.

Bastani, H., et al. (2025). *Generative AI without guardrails can harm learning*. *Proceedings of the National Academy of Sciences*. Source document.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). *The crowdless future? Generative AI and creative problem solving*. *Organization Science*. Source document.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). *Generative AI at work*. *The Quarterly Journal of Economics*, 140(2), 889-942. Source document.

Danry, V., et al. (2023). *Don't just tell me, ask me: AI systems that intelligently frame explanations as questions improve human logical discernment accuracy over causal AI explanations*. CHI 2023. Source document.

Danry, V., et al. (2024). *Deceptive AI systems that give explanations are more convincing than honest AI systems and can amplify belief in false information*. arXiv preprint. Source document.

Dell'Acqua, F., McFowland, E., Mollick, E. R., et al. (2026). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality*. *Organization Science*. Source document.

Doshi, A. R., & Hauser, O. P. (2024). *Generative AI enhances individual creativity but reduces the collective diversity of novel content*. *Science Advances*. Source document.

Drosos, I., et al. (2025). *It makes you think: Provocations help restore critical thinking to AI-assisted knowledge work*. arXiv preprint. Source document.

Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. (2023). *Ideas are dimes a dozen: Large language models for idea generation in innovation*. SSRN working paper. Source document.

Hwang, A. H.-C., et al. (2025). *It was 80% me, 20% AI: Seeking authenticity in co-writing with large language models*. *Proceedings of the ACM on Human-Computer Interaction*. Source document.

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., & Naaman, M. (2023). *Co-writing with opinionated language models affects users' views*. CHI 2023. Source document.

Kestin, G., et al. (2025). *AI tutoring outperforms in-class active learning*. *Scientific Reports*. Source document.

Klingbeil, A., Grutzner, L., & Schreck, P. (2024). *Trust and reliance on AI - An experimental study on the extent and costs of overreliance on AI*. *Computers in Human Behavior*, 160, 108352. Source document.

Kollock, P. (1998). *Social dilemmas: The anatomy of cooperation*. *Annual Review of Sociology*, 24, 183-214. Source document.

Lee, H.-P., Sarkar, A., Tankelevitch, L., et al. (2025). *The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers*. CHI 2025. Source document.

Lee, M., Liang, P., & Yang, Q. (2022). *CoAuthor: Designing a human-AI collaborative writing dataset for exploring language model capabilities*. CHI 2022. Source document.

Moon, K., et al. (2025). *Homogenizing effect of large language models on creative diversity: An empirical comparison of human and ChatGPT writing*. Source document.

Noy, S., & Zhang, W. (2023). *Experimental evidence on the productivity effects of generative artificial intelligence*. *Science*, 381(6654), 187-192. Source document.

Parasuraman, R., & Manzey, D. H. (2010). *Complacency and bias in human use of automation: An attentional integration*. *Human Factors*, 52(3), 381-410. Source document.

Peters, U., & Chin-Yee, B. (2025). *Generalization bias in large language model summarization of scientific research*. *Royal Society Open Science*. Source document.

Risko, E. F., & Gilbert, S. J. (2016). *Cognitive offloading*. *Trends in Cognitive Sciences*, 20(9), 676-688. Source document.

Shaer, O., et al. (2024). *AI-augmented brainwriting: Investigating the use of LLMs in group ideation*. arXiv preprint. Source document.

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. Source document.

Wei, L., Yu, H., Chen, Q., et al. (2025). *The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study*. *International Journal of Educational Technology in Higher Education*, 22, 23. Source document.