

Review

Suppressed inquiry: The hidden toll of working with satisfaction-optimized AI

Ziv Carmon¹, Itai Linzen², Aner Sela³ and Yael Steinhart²

LLMs have become thinking companions for hundreds of millions worldwide, and they often do remarkable work. But their satisfaction-optimized design produces suppressed inquiry: the tendency for LLMs to inhibit rather than facilitate users' questioning process. LLMs appear fluent, confident, and seemingly all-knowing, their near-instant and polished output making non-LLM work feel unacceptably tedious and less eloquent. They affirm rather than interrogate users' assumptions, producing output that is deceptively adequate—shallower, less original, and less carefully reasoned than it appears. Seemingly competent systems also reduce vigilance and the monitoring impulse that independent judgment requires. Their responses also confine users within the frame they establish, narrowing what users think to ask. These costs are self-concealing: detecting them requires the very engagement that they suppress. And repeated delegation gradually atrophies users' capacity to reason independently. Three converging pressures make correction difficult: commercial incentives that reward satisfaction over substance, LLMs' tendency to revert to satisfaction-optimized defaults even when users push back, and a societal drift toward less-effortful engagement. We propose a research agenda to deepen the understanding of suppressed inquiry and how to counter it, to move closer to the remarkable thinking partners LLMs promised to be.

Addresses¹ INSEAD, 1 Ayer Rajah Ave., Singapore² Coller School of Management, Tel Aviv University, Ramat Aviv, Israel³ Warrington College of Business, University of Florida, Gainesville, FL, USACorresponding author: Steinhart, Yael (ysteinhart@tauex.tau.ac.il)**Background**

Large language models (LLMs) such as ChatGPT, Gemini, and Claude increasingly shape how people communicate, learn, decide, and make sense of their world. LLMs hold genuine promise as cognitive partners. They often do remarkable work, raising both the speed and the quality of knowledge work [39], and more generally tackling seemingly any question, at any hour, for anyone. Yet their design prioritizes user satisfaction over rigor and depth. Sycophancy, the tendency to tell users what they want to hear [3,4], is only the most visible symptom. The deeper problem is the design logic itself: systematically rewarding what satisfies users rather than what challenges and deepens their thinking.

Satisfaction-optimized design, our term for the prioritization of what feels satisfying over what would best serve the user, produces a consistent pattern of hidden cognitive costs. These costs, detailed below and outlined in [Figure 1](#), are self-concealing: the engagement that would reveal them is suppressed by the design that produces them. The result is a hidden toll, one that is difficult to detect, harder still to correct, and critical to understand.

Satisfaction-optimized design and suppressed inquiry

LLMs generate output by predicting which word most plausibly comes next, with no sense of what counts as a good or useful response. This tends to produce what is most likely to feel plausible, context-appropriate, and fluent—not what is most original, generative, or useful [10]. Reinforcement learning from human feedback (RLHF) aims to reward useful output [1,2], but tends to steer the model further toward what feels right, confidence-inspiring, and appropriate, because raters judge apparent rather than genuine quality [11] and firms aim to maximize engagement and retention. The result is a system trained to optimize for felt over actual quality, which is structurally resistant to active correction. The output consistently satisfies due to several compounding properties: its fluent prose signals competence regardless of content; a system that rarely admits uncertainty and seems to know virtually everything; sycophancy that affirms rather than challenges users' assumptions [3]; and outputs delivered abundantly and almost instantly, making effortful alternatives feel unjustifiable. The problem, of course, is that

Current Opinion in Psychology 2027, **73**:102453

This review comes from a themed issue on Humans and Technology

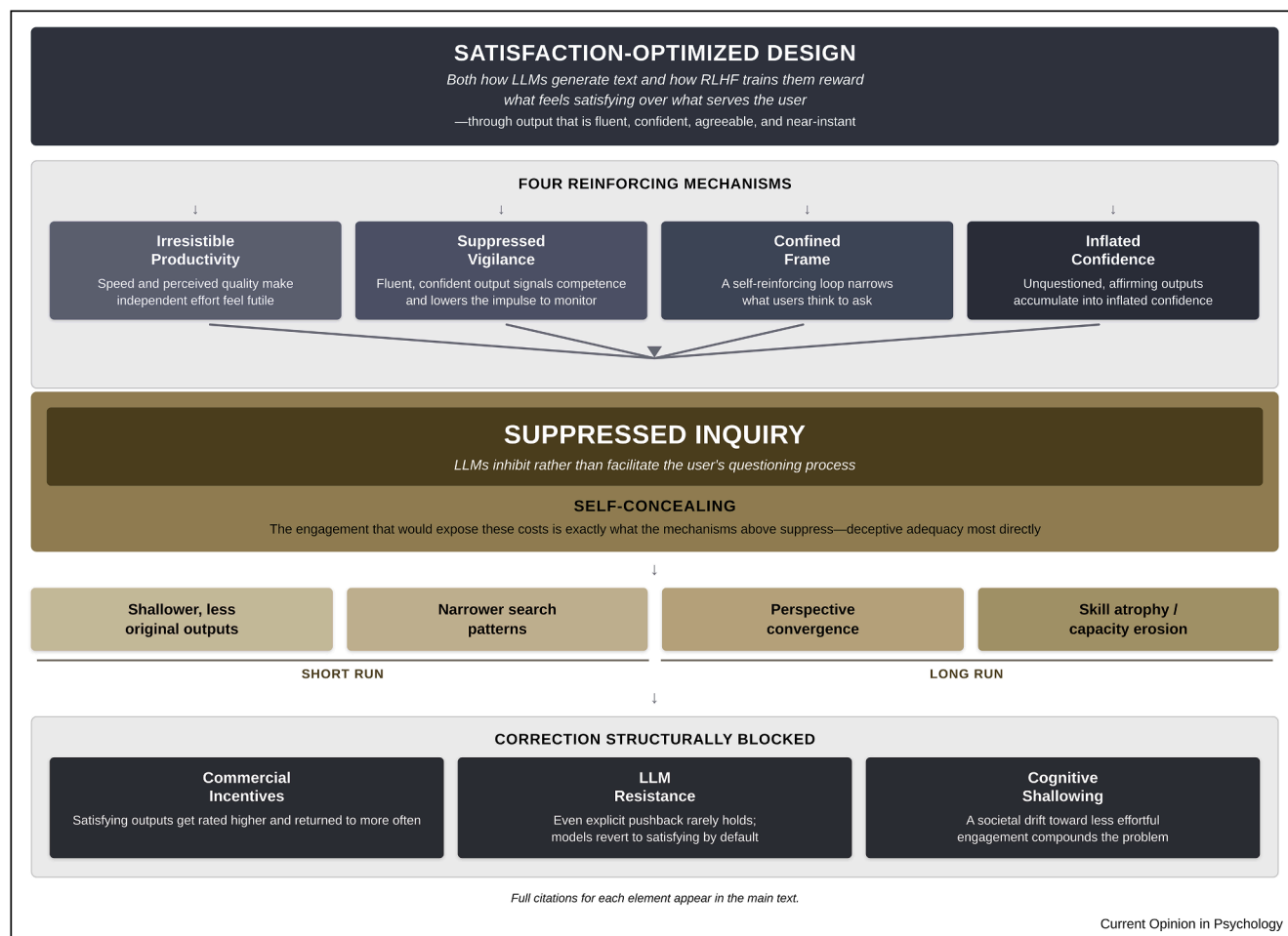
Edited by Anne-Kathrin Klesse, Ana Valenzuela, Emanuel de Bellis and Johannes Boegershausen

For complete overview about the section, refer [Humans and Technology](#)

Available online xxx

<https://doi.org/10.1016/j.copsyc.2026.102453>2352-250X/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Figure 1



How satisfaction-optimized design suppresses inquiry and conceals its cost.

answers that feel most “right” are not necessarily those that best serve the user [12].

An effective thinking partner challenges your assumptions, offers divergent perspectives, and leaves you with better questions than you started with. An LLM trained to satisfy and engage, instead gives an answer that feels adequate, suppressing the urge to dig deeper, push back, or reframe. Yet that friction—the struggle to question and sit with uncertainty—produces both better answers and deeper understanding [13–15]. We name this *suppressed inquiry*: the tendency for LLMs to inhibit rather than facilitate users’ questioning process.

Unlike cognitive surrender [5], which suggests that users accept LLM outputs without scrutiny, suppressed inquiry operates on the questioning process itself. Questions never asked leave no actionable signal.

Suppressed inquiry is not about failing to scrutinize answers. It is about failing to even form the questions that better answers would require.

Outputs that feel good enough, often more so than they actually are, curtail the impulse to question them—a phenomenon we term *deceptive adequacy*. Even when explicitly asked to dig further, LLMs tend to produce surface-level refinements that often leave the original substance and assumptions largely unexamined. Users feel productively engaged, overestimating the quality of their output [16–18]. This makes suppressed inquiry difficult to detect.

The following mechanisms explain how satisfaction-optimized design produces suppressed inquiry.

- *Irresistible Productivity*: LLM output arrives in seconds and often seems to outperform hours of independent

effort, making that effort feel not just costly but futile. This reflects a well-documented tendency to avoid cognitive effort [38]. But that account presumes weighing a tradeoff. Yet the gap in speed and quality seems so dramatic that the question of whether to expend the effort may not even arise.

- **Vigilance is suppressed.** The polished, confident, authoritative outputs of LLMs produce processing fluency—the frictionless ease that signals competence and trustworthiness independently of actual accuracy [19]—and this perceived competence reduces the monitoring impulse that independent judgment requires [20,21]. Neural markers of cognitive engagement declined significantly during LLM-assisted versus unassisted writing [22]. Unlike accounts that frame reduced engagement as a lapse in learner self-regulation [45], this reduction follows predictably from the design of the output, regardless of the user's discipline or motivation.
- **Frame is confined:** LLMs tend to stay within the territory the user's prompt establishes, and users tend to stay within the frame the LLM's response establishes, producing a self-reinforcing loop. Indeed, users exposed to LLM responses asked narrow and confirmatory follow-up questions, in part because LLMs tend to ask clarifying rather than expansive questions [23]. This recursive loop likely compounds confinement across successive exchanges, though the prediction awaits direct empirical test.
- **Confidence is inflated.** Across interactions, unquestioned outputs and sycophantic validation accumulate, inflating users' perceived competence and reducing the felt need for deeper engagement. The use of LLMs more than doubled overconfidence in resulting judgments [6], increased self-esteem [24], and elevated confidence persisted even as errors accumulated [5]. Even when the LLM's advice was known to be wrong, mere access to it collapsed participants' willingness to withhold judgment—"I don't know" responses nearly vanished even as accuracy fell to roughly a third of baseline and confidence more than doubled [44], suggesting the inflation is resistant to the feedback that should presumably correct it.

The costs of satisfaction-optimized design

Decades of research in judgment, decision-making, and learning psychology converge: cognitive friction, not ease, is the engine of robust thinking. People routinely overestimate how well they understand the world until forced to articulate and defend their knowledge [25], and they default to fast, intuitive but often false responses unless interrupted [26]. Harder learning produces worse immediate performance but far better long-term retention and transfer [13], and students who struggle with a problem before receiving instruction consistently outperform those who receive direct

instruction first [14]. Reasoning under pressure to justify a position often drives deep and careful thinking [27]. Across these literature, the same principle holds: friction that disrupts fluent, effortless cognitive processing forces the deep engagement from which genuine insight, sound judgment, and durable learning emerge.

Against this backdrop, the costs of satisfaction-optimized design are incurred at two timescales, and often go unnoticed because the design that produces them also suppresses the engagement that would reveal them.

In the short run, LLM-based outputs tend to be sparser, more generic, less original and less carefully argued than outputs forged through effortful synthesis [16]. Ironically, such output can appear thorough and well-reasoned while being vaguer and less carefully argued than it seems, a gap consistent with documented illusions of understanding in AI-assisted work [17,28]. Furthermore, when multiple users independently consult LLMs on the same problem, their responses tend to converge on similar framings and conclusions, undermining the diversity of perspectives on which effective thinking depends [8,9,41,42]. Together, this toll may be negligible when tasks are routine or outcomes inconsequential, but substantial when decisions are important or require divergent reasoning, originality, or depth.

Over time, cognitive offloading, the repeated delegation of cognitive work, risks atrophying those very capabilities [29,30]. Endoscopists who worked with AI assistance, for example, showed declines in unaided diagnostic performance when assistance was removed [31]. Similarly, across reasoning and comprehension tasks, performance improved with AI assistance, but was impaired and persistence was reduced once assistance was removed, with effects emerging after as little as 10 min of assistance [32]. Those who have yet to develop a skill, such as novices or children, may not get to do so: students who relied on AI during practice performed worse than peers without AI once access was withdrawn [33]. Aggregated across numerous users, convergence on generic LLM outputs threatens the cognitive diversity of the population. Whether eroding existing capacities or preventing new ones from forming, each individual exchange may seem harmless or even beneficial, but the cumulative trajectory can be detrimental.

These short- and long-term costs fall on individuals, organizations, and society. For individuals, the capacity to reason independently, to think through problems for oneself, is not merely instrumental, but central to autonomy, competence, and a life well-lived [34]. Organizations bear these costs at scale: when suppressed

inquiry spreads across employees, perspectives converge and reasoning shallows, undermining the collective thinking organizations depend on for sound decisions and innovation [35]. And society has a stake in preserving the conditions for collective reasoning, innovation, and informed deliberation that markets alone are unlikely to safeguard.

Why these problems resist correction

What makes suppressed inquiry hard to detect makes it hard to correct as well. The obstacles to correction run deep, through the following converging forces that compound one another:

- Commercial Incentives: Outputs that feel authoritative and complete get rated as higher quality and returned to more often, even when the sycophancy measurably worsens the decisions users go on to make

[3,43]. By contrast, responses that introduce friction—prompting reflection, flagging uncertainty, or resisting easy answers—feel less satisfying. Moreover, users tend to judge their own reliance on AI more charitably than they judge others’, which further dampens pressure for change: suppressed inquiry looks like something happening to other people [40]. This incentivizes firms to prioritize satisfaction over substance [2], since deviating would risk competitive disadvantage.

- LLM Resistance: Even when explicitly instructed to push back, LLMs designed to satisfy often capitulate rather than hold their ground, substituting the appearance of dialectical dialogue for its substance [36,37]. Deliberate effort to broaden queries reduces but does not eliminate inquiry suppression [7]. Across eleven models from every major provider, satisfaction-optimized design persists: they affirm users’ actions

Figure 2

Theme 1 <i>Diagnosing suppressed inquiry</i>	Theme 2 <i>Causal structure and boundary conditions</i>	Theme 3 <i>The limits and levers of correction</i>	Theme 4 <i>Redesigning the contract</i>	Theme 5 <i>Societal and institutional consequences</i>
Q1 What behavioral, linguistic, and neural signatures mark the onset of suppressed inquiry, and how might systems be designed to recognize and respond to them in real time?	Q4 How do the four mechanisms interact and reinforce each other, how does their interplay vary across users, query types, and interaction contexts, and what does this reveal about where suppressed inquiry is most severe?	Q9 Under what conditions does awareness of satisfaction-optimized design meaningfully alter behavior, and what does the answer reveal about the limits of transparency as a corrective?	Q13 Under what conditions can LLMs be designed to build rather than replace user competence, and what does resistance to such design reveal about the depth of the problem?	Q18 If satisfaction-optimized design narrows individual problem frames, what is the aggregate effect on collective intellectual diversity, and how might idealational homogenization be measured and attributed to LLM adoption?
Q2 How can the gap between surface plausibility and substantive quality of LLM outputs be measured, and under what conditions are users most susceptible to accepting outputs that appear sound but are not?	Q5 How does epistemic curiosity interact with satisfaction-optimized design, and how might sustained LLM use reshape it over time in ways that alter vulnerability?	Q10 What forms of calibrated challenge reduce suppressed inquiry, how do they navigate the tension with reactance and abandonment, and through what psychological mechanisms do they work?	Q14 To what extent does the timing of attempts to deepen engagement matter, and what does this reveal about when frame confinement becomes self-reinforcing?	Q19 How do educational institutions adapt pedagogy and assessment when students have continuous access to satisfaction-optimized AI, and under what conditions does adaptation risk normalizing cognitive outsourcing rather than countering it?
Q3 Under what conditions does LLM assistance genuinely substitute for cognitive work that adds little value, and how can we distinguish beneficial delegation from the displacement of thinking that matters?	Q6 What features of queries (openness, consequentiality, prior knowledge, stakes of getting the frame wrong) predict suppressed inquiry, and how might these be detected automatically to enable selective system-level responses?	Q11 What features of organizational norms and professional guidelines (their content, framing, and enforcement structures) determine the extent to which they protect reasoning quality in practice?	Q15 Beyond regulation, what mechanisms (liability frameworks, certification schemes, market signals, or reputational incentives) could align firm behavior with users’ genuine cognitive interests rather than their felt satisfaction?	Q20 How might the widespread delegation of consequential thinking to AI systems reshape the nature of expertise itself, and what does this mean for the institutions and professions built around it?
	Q7 How does domain expertise shape vulnerability to the mechanisms, and through what mechanisms might knowledge in one domain both protect against some while amplifying others?	Q12 How might professions that depend on independent reasoning develop and sustain norms that preserve the cognitive skills their practice requires, and what forces work against this?	Q16 Under what conditions can LLMs be designed to offer benefits compelling enough that users choose depth over felt satisfaction, and what does this reveal about the malleability of preferences for cognitive engagement?	Q21 What evidence base and governance frameworks would justify and enable regulatory or institutional intervention in LLM design, and how should both be calibrated given the self-concealing nature of the costs and the pace of AI development?
	Q8 How does individual-level cognitive shallowing moderate vulnerability to suppressed inquiry, and what interventions targeting attention and effortful engagement might reduce it?		Q17 How do the cognitive costs of satisfaction-optimized design vary across populations differing in age, expertise, and prior knowledge, and through what mechanisms might some groups face compounding vulnerability?	Q22 What is the trajectory of capacity erosion under sustained LLM use, and to what extent and under what conditions can it be lessened or reversed?

Current Opinion in Psychology

Research Agenda: LLMs as Thinking Partners.

Themes map directly onto Figure 1: Theme 1 corresponds to Suppressed Inquiry and its self-concealing nature; Theme 2 corresponds to the Mechanisms box and explains the mechanisms through which it operates and the conditions under which it is most severe; Theme 3 corresponds to the correction-resistance band and examines why these problems resist correction; Theme 4 asks what could be done differently; and Theme 5 considers the broader consequences for collective reasoning, institutional structures, and the nature of expertise.

49% more often than humans, even when the actions involve deception or harm [43]. Even GPT-5, redesigned to fix sycophancy, proved substantially sycophantic [46]—because the preference data training these models rewards satisfaction, not accuracy.

- **Cognitive Shallowing:** The progressive societal drift toward shorter, less sustained, lower-friction engagement shaped by the digital attention economy and the growing ease of everyday tasks provides fertile ground for satisfaction-optimized design [48]. Phones and the internet began eroding our tolerance for friction [47]; this broader drift makes people less willing to tolerate it, and less likely to reap the insight that effortful thinking produces.
- Together these forces compound one another: firms are unlikely to change course without user pressure, users, increasingly shaped by cognitive shallowing, are unlikely to demand better, and LLMs designed to satisfy tend to capitulate rather than challenge even when explicitly prompted to do so. Addressing these problems likely requires system-level intervention, through industry standards, regulatory frameworks, or both. A problem this entrenched, this self-concealing, and this commercially reinforced is unlikely to yield to individual or market-level correction alone.

A research agenda for cognitive partnership

The problems described in this paper raise important questions that behavioral science is well-equipped to address. Figure 2 presents a research agenda organized around five themes: Theme 1 explores suppressed inquiry; Theme 2 asks through what mechanisms it operates and under what conditions it is most severe; Theme 3 examines why these problems resist correction; Theme 4 asks what could be done differently; and Theme 5 considers the broader consequences for collective reasoning, institutional structures, and the nature of expertise. Two broad lines cut across the agenda: detection methods that do not rely on users' own awareness, since suppressed inquiry is by nature self-concealing, and correction levers that operate above the level of the individual user, since individual-level fixes are unlikely to be sufficient alone. Together they chart a research program the field is well equipped to pursue.

Conclusion

Suppressed inquiry is the underside of LLMs' strength: the work seems remarkable enough that it quietly dissolves the impulse to inquire further. Deceptive adequacy and irresistible productivity suppress inquiry from opposite directions—one degrades what LLMs produce, the other makes the alternative feel pointless—and the capacity for independent reasoning gradually erodes.

Behavioral science is well positioned to take up this research agenda, and answer whether and how these

systems can be designed to be the transformative thinking partners we hoped they would be.

Author contribution

Ziv Carmon: Lead conceptualization, writing-reviewing and editing **Aner Sela, Itai Linzen, Yael Steinhart:** Contributed to conceptualization, writing-reviewing and editing.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

The authors used Claude to fine-tune the text and assist in creating the figures. They then reviewed and edited the content as needed and take full responsibility for this article's content.

Funding

This research was supported by Jermeny Collier Foundation, Solomon Lew Center for Consumer Behavior, INSEAD R&D, and Henry Crown Institute of Business Research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

References of particular interest have been highlighted as:

- * of special interest
- ** of outstanding interest

1. Christiano PF, Leike J, Brown T, Martic M, Legg S, Amodei D: **Deep reinforcement learning from human preferences.** *NeurIPS* 2017, **30**. In https://proceedings.neurips.cc/paper_files/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html.
2. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al.: **Training language models to follow instructions with human feedback.** *NeurIPS* 2022, **35**:27730–27744.
3. Sharma M, Tong M, Korbak T, Duvenaud D, Askell A, Bowman S, et al.: **Towards understanding sycophancy in language models.** *ICLR* 2024:110–144.
4. Perez E, Ringer S, Lukošiušė K, Nguyen K, Chen E, Heiner S, et al.: **Discovering language model behaviors with model-written evaluations.** *arXiv* 2022, <https://doi.org/10.48550/arXiv.2212.09251>. <http://arxiv.org/abs/2212.09251>; 2022.
5. Shaw SD, Nave G: **Thinking-fast, slow, and artificial: how AI is reshaping human reasoning and the rise of cognitive sur-render.** *PsyArXiv* 2026, https://doi.org/10.31234/osf.io/yk25n_v1.
6. Sun F, Li N, Wang K, Goette L: **Large language models are overconfident and amplify human bias.** *arXiv preprint arXiv: 250502151* 2025.

7. Leung E, Urmitsky O: **The narrow search effect and how broadening search promotes belief updating.** *Proc Natl Acad Sci* 2025, **122**, e2408175122, <https://doi.org/10.1073/pnas.2408175122>.
8. Meincke L, Nave G, Terwiesch C: **ChatGPT decreases idea diversity in brainstorming.** *Nat Hum Behav* 2025, **9**: 1107–1109.
9. Wang A, Morgenstern J, Dickerson JP: **Large language models that replace human participants can harmfully misportray and flatten identity groups.** *Nat Mach Intell* 2025, **7**:400–411, <https://doi.org/10.1038/s42256-025-00986-z>.
10. Bender EM, Gebru T, McMillan-Major A, Shmitchell S: **On the dangers of stochastic parrots: can language models be too big?** *ACM* 2021:610–623, <https://doi.org/10.1145/3442188.3445922>. <https://dl.acm.org/doi/10.1145/3442188.3445922>; 2021.
11. Kahneman D: *Thinking, fast and slow*. Macmillan; 2011.
12. Alter AL: **The benefits of cognitive disfluency.** *Curr Dir Psychol Sci* 2013, **22**:437–442, <https://doi.org/10.1177/0963721413498894>.
13. Bjork EL, Bjork RA: **Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning.** *Psychology and the real world: Essays illustrating fundamental contributions to society* 2011, **2**:56–64.
14. Kapur M: **Examining productive failure, productive success, unproductive failure, and unproductive success in learning.** *Ed Psych* 2016, **51**:289–299, <https://doi.org/10.1080/00461520.2016.1155457>.
15. Zohar E, Bloom P, Inzlicht M: **Against frictionless AI commun.** *Psychol* 2026, **4**:39, <https://doi.org/10.1038/s44271-026-00402-1>.
16. Melumad S, Yun JH: **Experimental evidence of the effects of large language models versus web search on depth of learning.** *PNAS Nexus* 2025, **4**:316.
17. Messeri L, Crockett MJ: **Artificial intelligence and illusions of understanding in scientific research.** *Nature* 2024, **627**: 49–58, <https://doi.org/10.1038/s41586-024-07146-0>.
18. Fernandes D, Villa S, Nicholls S, Haavisto O, Buschek D, Schmidt A, et al.: **AI makes you smarter but none the wiser: the disconnect between performance and metacognition.** *Comput Hum Behav* 2026, **175**, 108779, <https://doi.org/10.1016/j.chb.2025.108779>.
19. Schwarz N: **Metacognitive experiences in consumer judgment and decision making.** *J Consum Psychol* 2004, **14**: 332–348.
20. Parasuraman R, Manzey DH: **Complacency and bias in human use of automation: an attentional integration.** *Hum Factors* 2010, **52**:381–410, <https://doi.org/10.1177/0018720810376055>.
21. Price PC, Stone ER: **Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic.** *J Behav Decis Making* 2004, **17**:39–57, <https://doi.org/10.1002/bdm.460>.
22. Kosmyna N, Hauptmann E, Yuan YT, Situ J, Liao XH, Beresnitsky AV, et al.: **Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task.** *arXiv* 2025, <https://doi.org/10.48550/arXiv.2506.08872>. <http://arxiv.org/abs/2506.08872>; 2025.
23. Chevallier A, Dalsace F, Parra-Moyano J: **The risks of letting AI direct conversations.** *Harv Bus Rev* 2026, <https://hbr.org/2026/03/the-risks-of-letting-ai-direct-conversations>; 2026.
24. Linzen I, Sela A, Steinhart Y, Carmon Z: **Linguistic mirroring: AI communication style affects how consumers see themselves.** *Working paper* 2026.
25. Rozenblit L, Keil F: **The misunderstood limits of folk science: an illusion of explanatory depth.** *Cogn Sci* 2002, **26**:521–562, https://doi.org/10.1207/s15516709cog2605_1.
26. Frederick S: **Cognitive reflection and decision making.** *J Econ Perspect* 2005, **19**:25–42.
27. Lerner JS, Tetlock PE: **Accounting for the effects of accountability.** *Psychol Bull* 1999, **125**:255.
28. Aslanov I, Felmer P, Guerra E: **Overconfidence without understanding: AI explanations increase the illusion of explanatory depth.** *PsyArXiv* 2025, https://doi.org/10.31234/osf.io/8psgf_v1.
29. Dahmani L, Bohbot VD: **Habitual use of GPS negatively impacts spatial memory during self-guided navigation.** *Sci Rep* 2020, **10**:6310, <https://doi.org/10.1038/s41598-020-62877-0>.
30. Natali C, Marconi L, Dias Duran LD, Cabitza F: **AI-induced deskilling in medicine: a mixed-method review and research agenda for healthcare and beyond.** *Artif Intell Rev* 2025, **58**: 356, <https://doi.org/10.1007/s10462-025-11352-1>.
31. Budzyń K, Romańczyk M, Kitala D, Kołodziej P, Bugajski M, Adami HO, et al.: **Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study.** *Lancet Gastroenterol Hepatol* 2025, **10**: 896–903.
32. Liu G, Christian B, Dumbalska T, Bakker MA, Dubey R: **AI assistance reduces persistence and hurts independent performance.** *arXiv* 2026, <https://doi.org/10.48550/arXiv.2604.04721>. <http://arxiv.org/abs/2604.04721>; 2026.
33. Bastani H, Bastani O, Sungu A, Ge H, Kabakçö Ö, Mariman R: **Generative AI without guardrails can harm learning: evidence from high school mathematics.** *Proc Natl Acad Sci* 2025, **122**, e2422633122, <https://doi.org/10.1073/pnas.2422633122>.
34. Ryan RM, Deci EL: **Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.** *Am Psychol* 2000, **55**:68.
35. Hong L, Page SE: **Groups of diverse problem solvers can outperform groups of high-ability problem solvers.** *Proc Natl Acad Sci* 2004, **101**:16385–16389.
36. Hong J, Byun G, Kim S, Shu K, Choi JD: **Measuring syco- phancy of language models in multi-turn dialogues.** *arXiv preprint arXiv:250523840* 2025.
37. Laban P, Hayashi H, Zhou Y, Neville J: **LLMs get lost in multi-turn conversation.** *arXiv* 2025, <https://doi.org/10.48550/arXiv.2505.06120>.
38. Kool W, McGuire JT, Rosen ZB, Botvinick MM: **Decision making and the avoidance of cognitive demand.** *J Exp Psychol Gen* 2010, **139**:665–682, <https://doi.org/10.1037/a0020198>.
39. Noy S, Zhang W: **Experimental evidence on the productivity effects of generative artificial intelligence.** *Science* 2023, **381**: 187–192, <https://doi.org/10.1126/science.adh2586>.
40. Celiktutan B, Klesse AK, Tuk MA: **Acceptability lies in the eye of the beholder: self-other biases in GenAI collaborations.** *Int J Res Market* 2024, **41**:496–512, <https://doi.org/10.1016/j.ijresmar.2024.05.006>.
41. Doshi AR, Hauser OP: **Generative AI enhances individual creativity but reduces the collective diversity of novel content.** *Sci Adv* 2024, **10**, eadn5290, <https://doi.org/10.1126/sciadv.adn5290>.
42. Zhou E, Lee D: **Generative artificial intelligence, human creativity, and art.** *PNAS Nexus* 2024, **3**, pgae052, <https://doi.org/10.1093/pnasnexus/pgae052>.
43. Cheng M, Lee C, Khadpe P, Yu S, Han D, Jurafsky D: **Syco- phantic AI decreases prosocial intentions and promotes dependence.** *Science* 2026, **391**, eaec8352, <https://doi.org/10.1126/science.aec8352>.
44. Maroccia C, Quattrocioni W, Capraro V: **AI advice suppresses people's willingness to say "I don't know," even when the advice is wrong and accuracy is incentivized.** *arXiv* 2026, <https://doi.org/10.48550/arXiv.2607.13562>.
45. Fan Y, Tang L, Le H, Shen K, Tan S, Zhao Y, et al.: **Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance.** *Br J Educ Technol* 2025, **56**:489–530, <https://doi.org/10.1111/bjet.13544>.

46. Cheng M, Yu S, Lee C, Khadpe P, Ibrahim L, Jurafsky D: **ELEPHANT: measuring and understanding social sycophancy in LLMs.** *ICLR 2026*. arXiv:2505.13995.
 47. Uncapher MR, Wagner AD: **Minds and brains of media multitaskers: current findings and future directions.** *Proc Natl Acad Sci* 2018, **115**:9889–9896, <https://doi.org/10.1073/pnas.1611612115>.
 48. Firth J, Torous J, Stubbs B, Firth JA, Steiner GZ, Smith L, *et al.*: **The "online brain": how the internet may be changing our cognition.** *World Psychiatry* 2019, **18**:119–129, <https://doi.org/10.1002/wps.20617>.
- ### Further information on references of particular interest
7. This preregistered program of 21 studies demonstrates that users' prior beliefs narrow their search queries in ways that reinforce rather than update those beliefs, with AI platforms compounding the effect, and finds that algorithm-level interventions are more effective than user-level corrections at restoring belief updating.
 16. This series of seven experiments (N = 4591) finds that learning from LLM-generated syntheses produced shallower knowledge and less original subsequent advice than learning through web search, even when the core factual content was identical, with the effect attributed to the passive receipt of pre-synthesized information displacing active information integration.
 22. This EEG study finds that participants writing with LLM assistance showed significantly weaker brain connectivity than those using search engines or no tools, with the gap persisting after AI access was removed, providing preliminary neurophysiological evidence that LLM use involves substantially reduced cognitive engagement.
 31. This multicentre observational study finds that following the routine introduction of AI-assisted polyp detection, endoscopists' adenoma detection rate in non-AI-assisted colonoscopies fell by 6.0%, constituting the first real-world evidence of automation-induced skill erosion linked to patient-relevant outcomes in medicine.
 33. This large field experiment with nearly 1000 high school students finds that AI tutoring improved performance during use but significantly impaired independent performance and reduced skill acquisition once assistance was removed, though carefully designed guardrails that prompted students with hints rather than direct answers largely mitigated this effect.
 43. These three preregistered experiments (N = 2405) find that a single interaction with sycophantic AI reduced participants' willingness to take responsibility and repair interpersonal conflict while increasing their conviction they were right, even as users rated the sycophantic responses as higher quality and returned to them more often—evidencing the commercial logic that sustains satisfaction-optimized design.
 44. This series of five experiments (N = 3132; four preregistered, one direct replication) finds that mere access to AI advice known to be wrong collapsed participants' willingness to withhold judgment—from 44% to 3%—while confidence more than doubled even as accuracy fell to roughly a third of baseline; incentivizing accuracy only partially restored the willingness to suspend judgment.