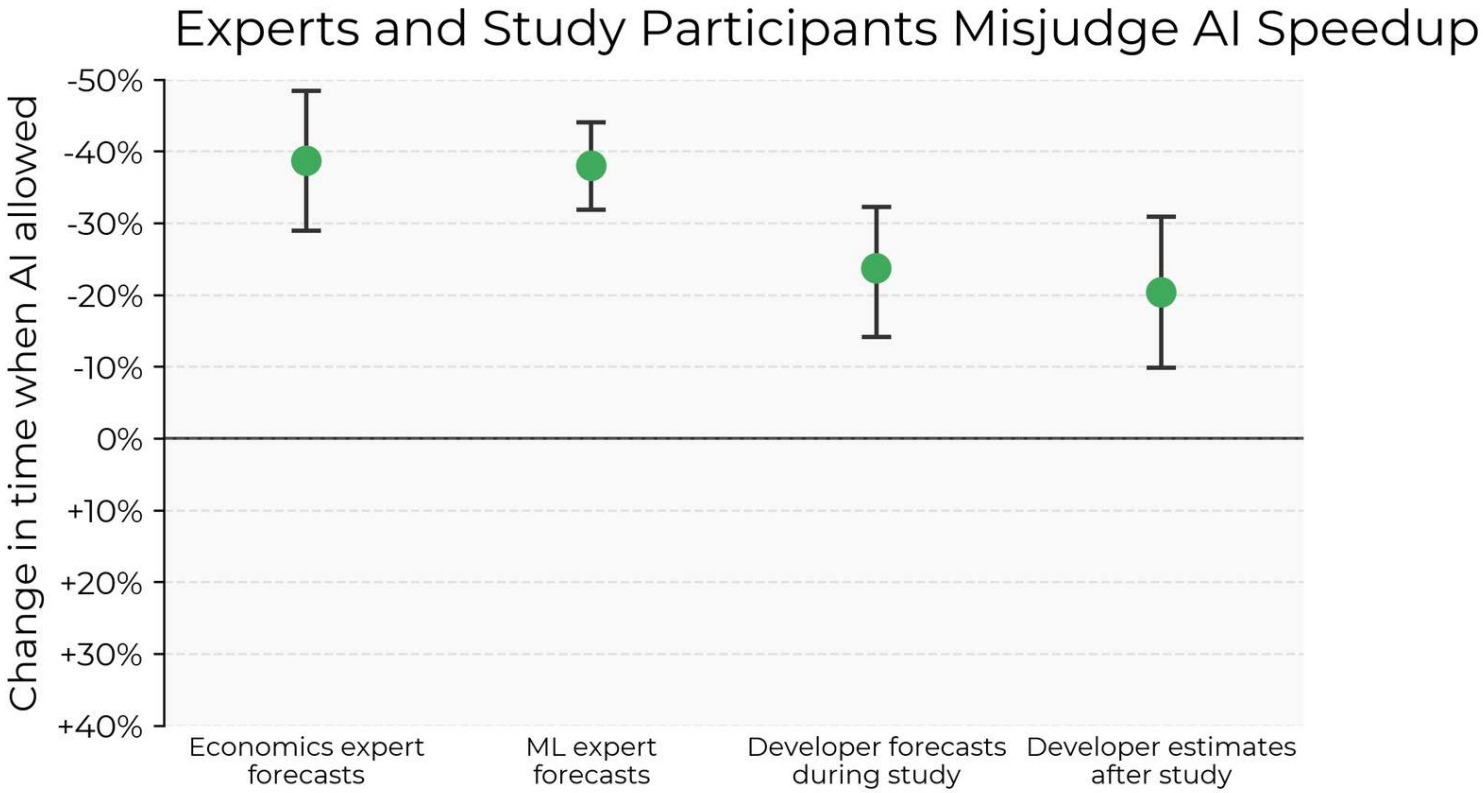
Human-AI Interaction, Lecture 17

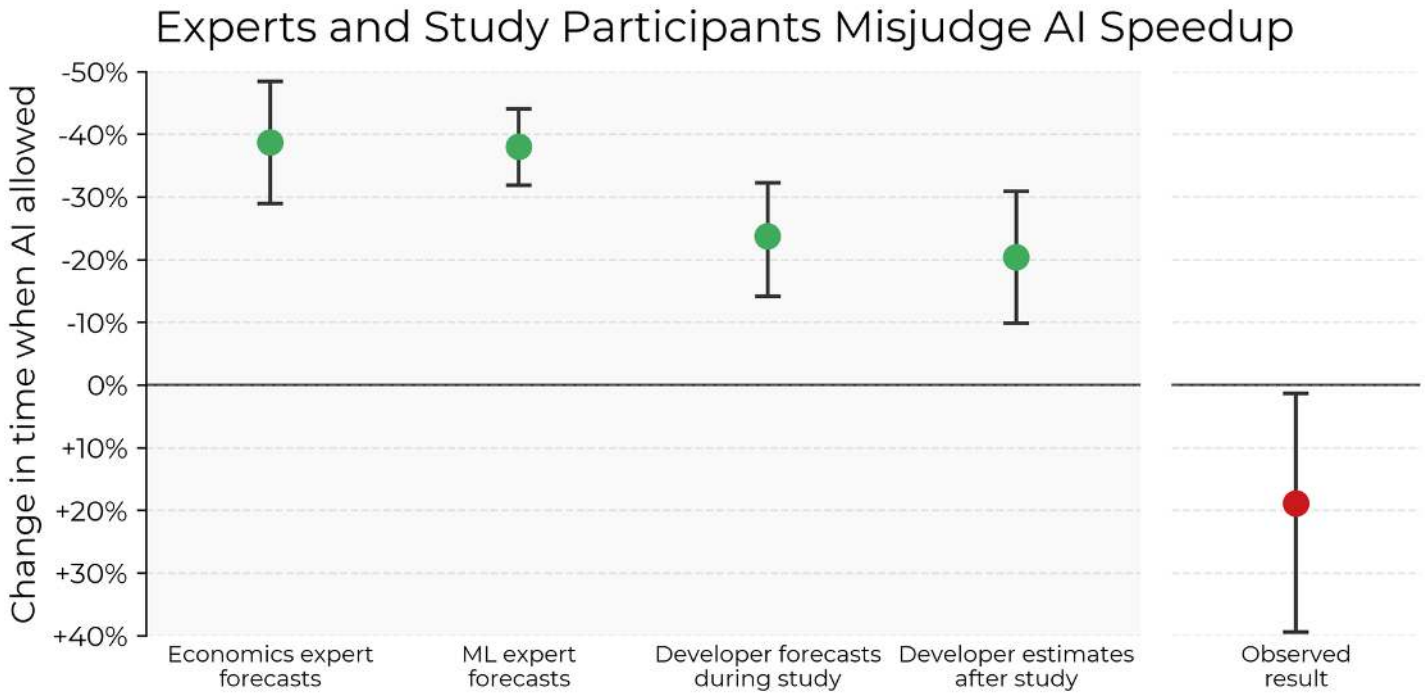
AI, Cognition, and Work

Forecasted speedup



Becker, Rush, Barnes & Rein (METR), 2025, Fig. 1

Measured result: **19% longer** with AI

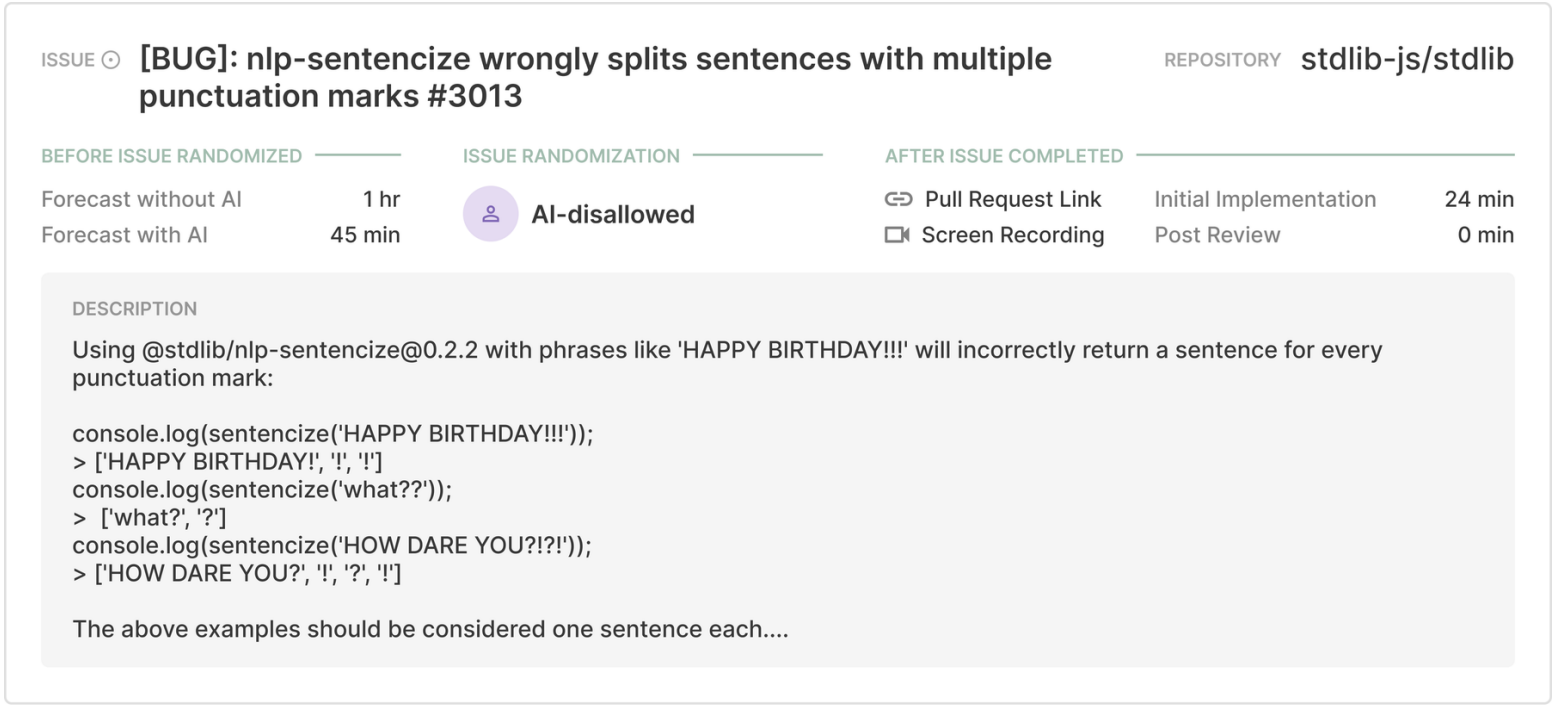


They still believed AI made them **20% faster**.

Becker et al., METR 2025, Fig. 1

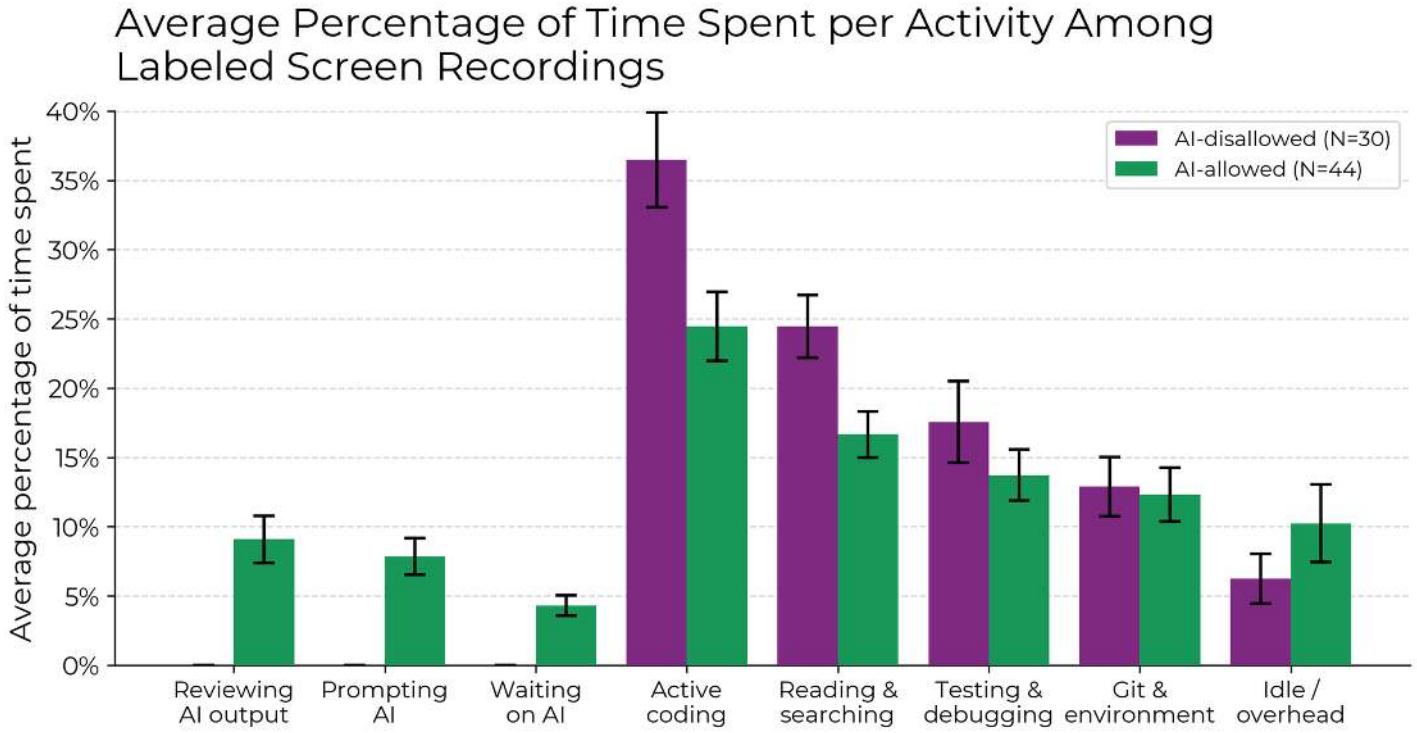
Study design: 16 developers, 246 issues

A coin flip decided whether AI was allowed.



Becker et al.,
METR 2025, Fig. 3

Where the developers' time went



They accepted less than 44% of the code the AI wrote.

Becker et al., METR 2025, Fig. 6 and Table 1

When AI does part of a job, what happens to the **person** who does the rest?

First reading: Bainbridge, “Ironies of Automation”

Irony: “combination of circumstances, the result of which is the direct **opposite** of what might be expected”

Bainbridge, Automatica 1983, p. 775

The tasks left over

The second irony is that the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate.

Bainbridge, Automatica 1983, p. 775

Loss of skills

Unfortunately, physical skills deteriorate when they are not used, particularly the refinements of gain and timing. This means that a formerly experienced operator who has been monitoring an automated process may now be an inexperienced one.

Bainbridge, Automatica 1983, p. 775

2013: the FAA's advice to airlines


U.S. Department
of Transportation
Federal Aviation
Administration

SAFO

Safety Alert for Operators

SAFO 13002
DATE: 1/4/13

Flight Standards Service

“Unfortunately, continuous use of those systems **does not reinforce** a pilot's knowledge and skills in manual flight operations.”

Federal Aviation Administration, Safety Alert for Operators 13002, “Manual Flight Operations,” 2013

Monitoring the automation

We know from many 'vigilance' studies (Mackworth, 1950) that it is impossible for even a highly motivated human being to maintain effective visual attention towards a source of information on which very little happens, for more than about half an hour.

Bainbridge, Automatica 1983, p. 776

"The human monitor has been given an **impossible task**."

Tempe, Arizona, March 18, 2018

“the better the automation system, the more likely the operator is to become **complacent** and not detect its failure”

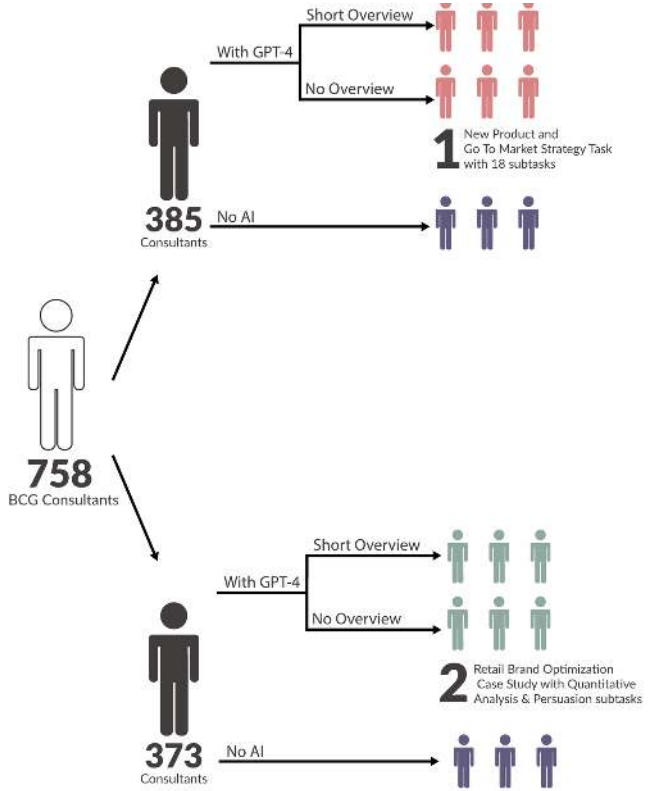


National Transportation Safety Board, Highway Accident Report NTSB/HAR-19/03, 2019, cover and p. 43

“By taking away the easy parts of his task, automation can make the difficult parts of the human operator's task **more difficult**.”

Bainbridge, Automatica 1983, p. 777

Experiments with 758 consultants at BCG



Note. This figure illustrates the experimental setup utilized in our study.

Second reading: Dell'Acqua et al., 2026

Three groups:

no AI

GPT-4

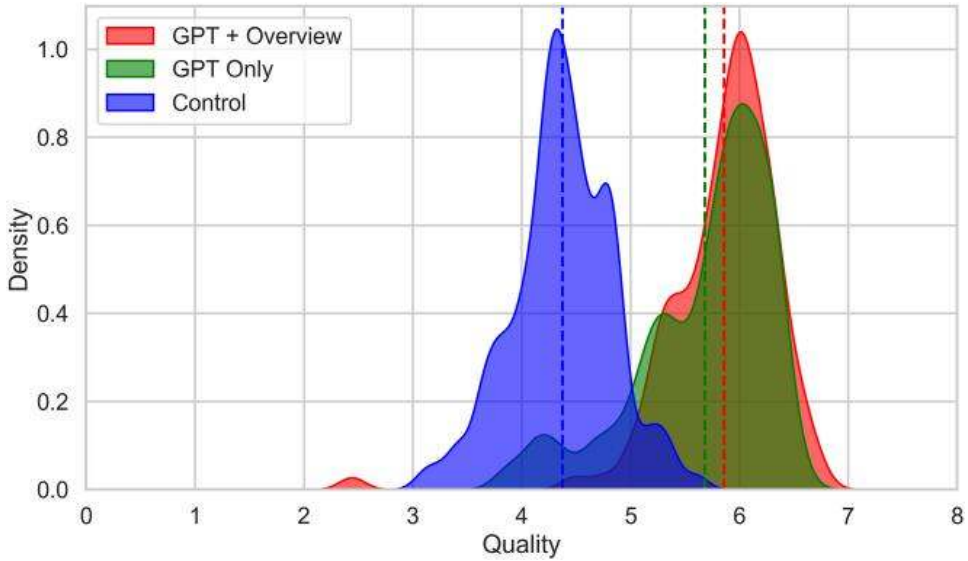
GPT-4 and a prompting overview

Dell'Acqua et al., Organization Science 2026, Fig. 2

Inside the frontier: a new shoe

creativity

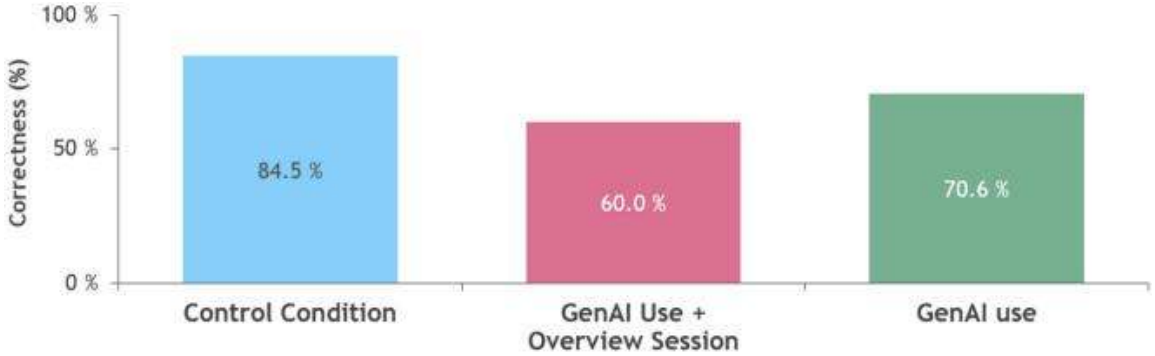
“Propose at least 10 ideas for a new shoe targeting an underserved market or sport.”



With GPT-4: **12.2% more tasks, 25.1% faster**

Dell'Acqua et al., Organization Science 2026, Section 4.1, abstract and Fig. 3

Outside the frontier: correct answers



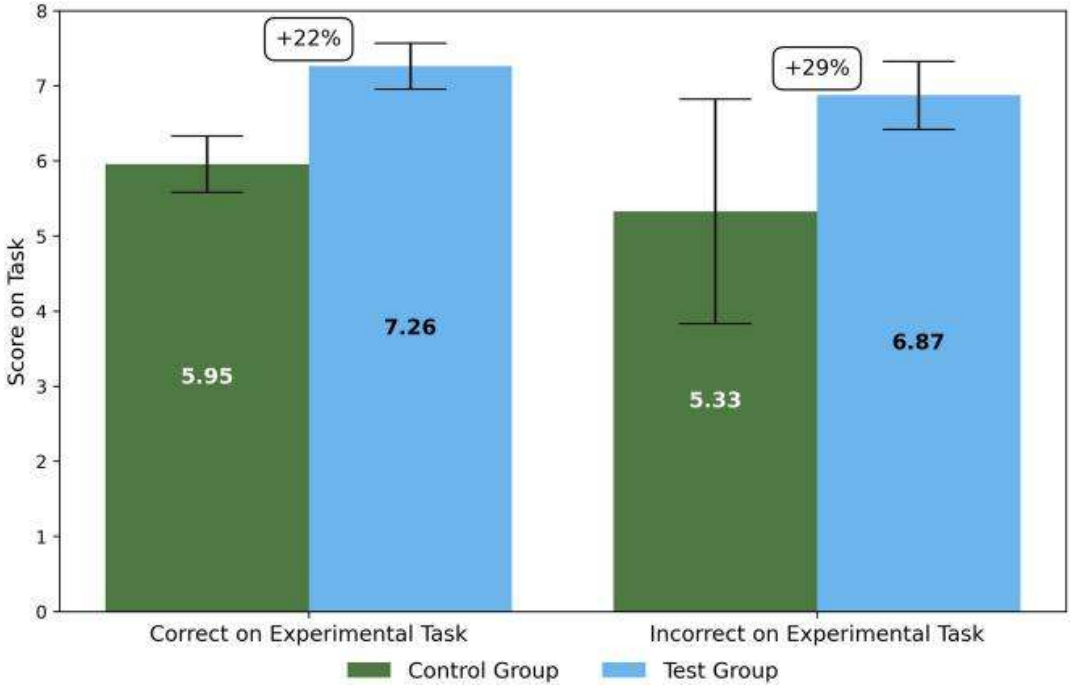
No AI: **84.5%**

GPT-4: **70.6%**

GPT-4 + overview: **60.0%**

Dell'Acqua et al., Organization Science 2026, Fig. 5 and Section 4.2

Persuasiveness of recommendations

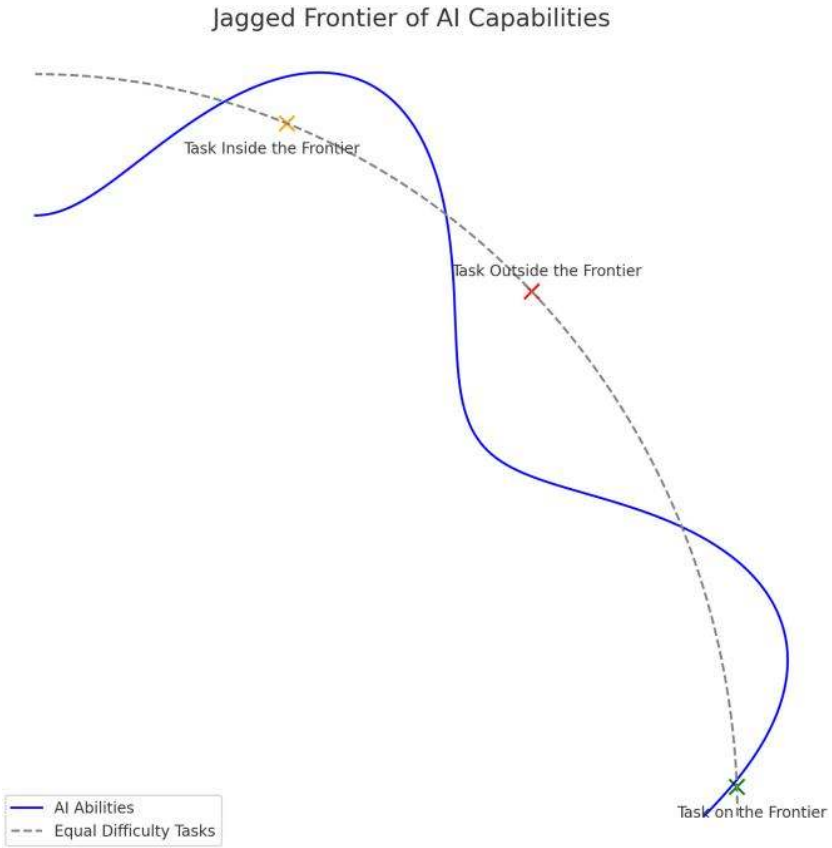


"AI improves presentation and argumentation even when the underlying analysis is **flawed**."

green: no AI
blue: with AI

Dell'Acqua et al., Organization Science 2026, Fig. 6 and Section 4.2

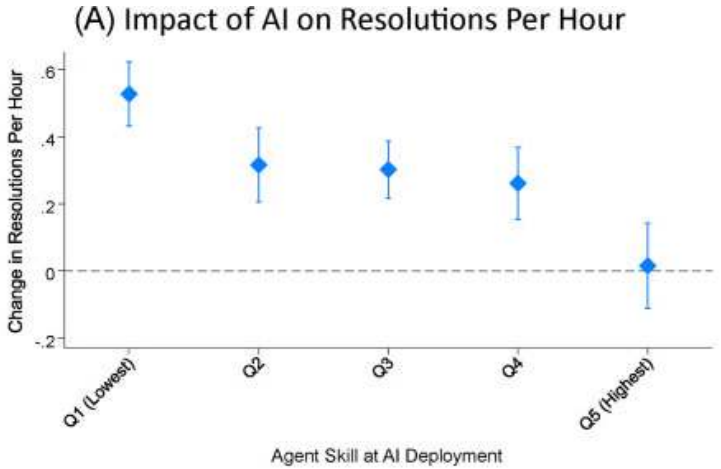
A **jagged** frontier



People who cannot tell which side a task is on can “perform worse as a result of becoming overly reliant on AI.”

Dell'Acqua et al., Organization Science 2026, Fig. 1 and Section 6

Customer support with AI



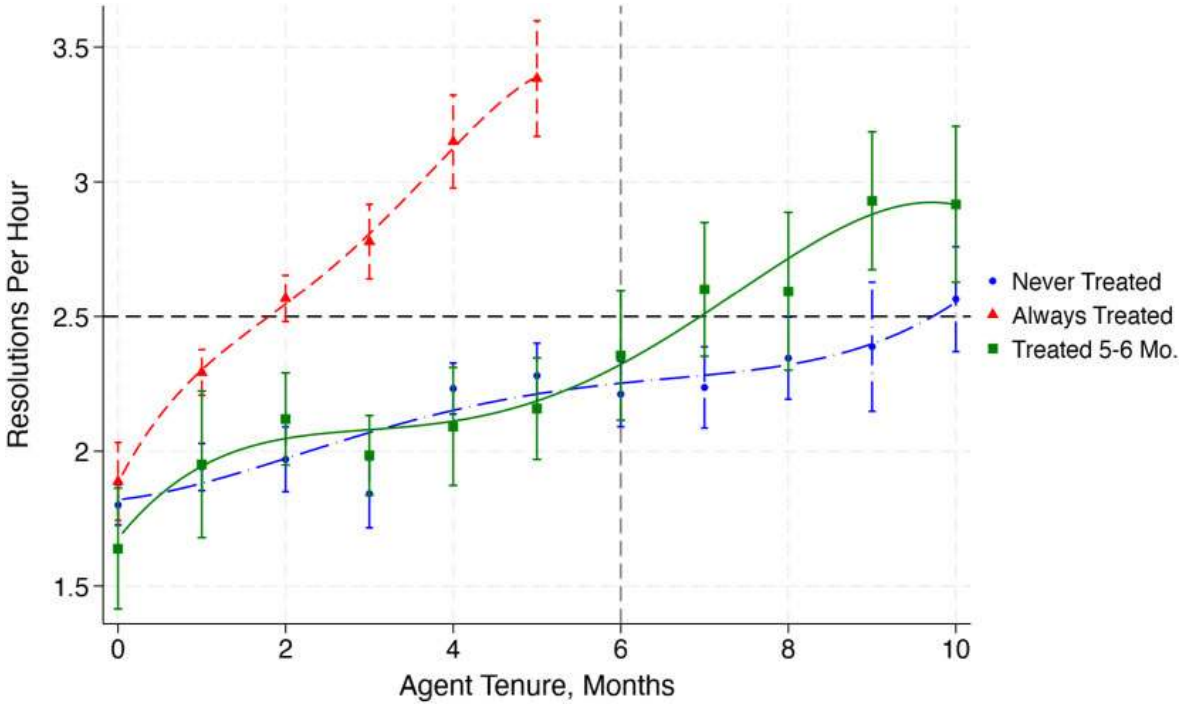
Brynjolfsson, Li & Raymond, QJE 2025, Fig. III(A) and Section V.B

Average: **+15%** issues resolved per hour

Lowest-skill fifth: **+36%**

Highest-skill fifth: no significant change

Learning curves of new agents



red: AI from the start

blue: never

green: AI from month 5

Brynjolfsson, Li & Raymond, QJE 2025, Fig. V and Section VI.B

Discussion: Bainbridge and Dell'Acqua

1. Which of Bainbridge's ironies do you see in the METR study?

2. How would you find out which side of the frontier a task in your own project is on?

3. If AI lets new workers skip the first six months, how does anyone become an **expert**?

Your week of AI use

1. **Alone, 4 min**: fill in the log for the last 7 days
2. **Pairs, 3 min**: what did you both stop doing?
3. **Room, 4 min**: we collect patterns on the board

Log of AI use

Day	What I used AI for	Tool	What I stopped doing myself	Did I check it? How?
Mon	*summarized a reading*	*ChatGPT*	*reading the whole paper*	*no*

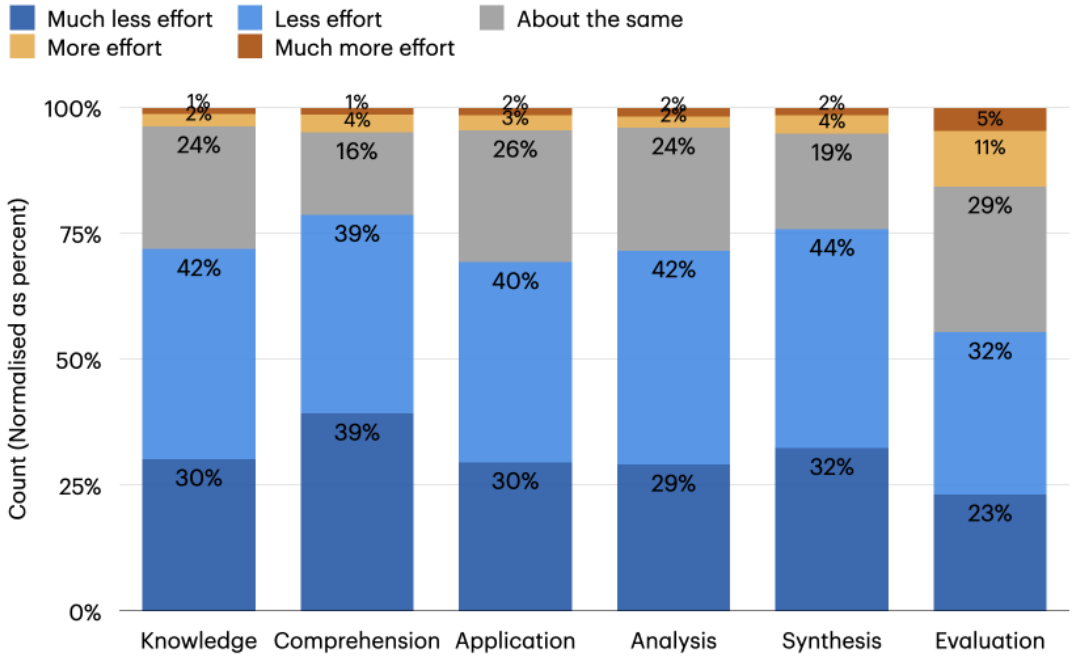
One skill I want to keep: _____

The grey row is an example.

Thinking with AI

Three studies on effort, opinions and memory

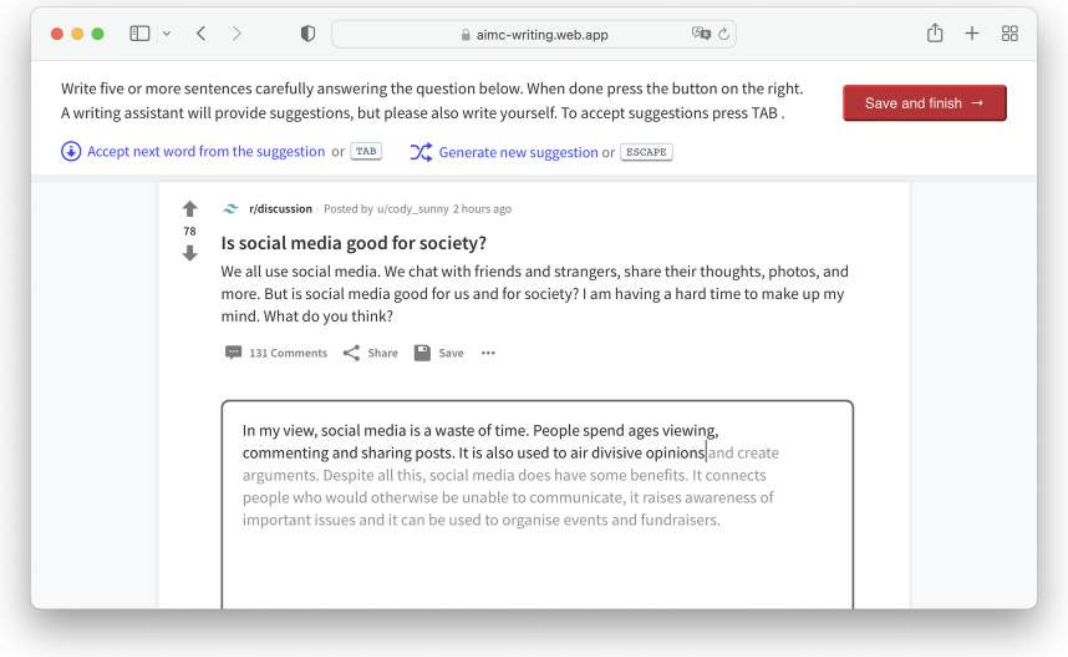
Effort in critical thinking



Lee et al., CHI 2025, Fig. 2, Section 5.1 and abstract

"higher confidence in GenAI is associated with **less critical thinking**"

An **opinionated** writing assistant



Both their posts and their survey answers afterward moved toward the model's side.

Jakesch, Bhat, Buschek, Zalmanson & Naaman, CHI 2023, Fig. 2, Sections 4.1 to 4.5

Optional reading: EEG while writing

Three groups: ChatGPT, a search engine, or no tools



Kosmyna et al., arXiv 2506.08872, Fig. 4

Brain connectivity and memory

Connectivity: strongest with no tools, weakest with ChatGPT

15 of 18 in the ChatGPT group could not correctly quote their own essay

Kosmyna et al., arXiv 2025, abstract and session 1 interviews (Question 3)

Limits of the study

Is it safe to say that LLMs are, in essence, making us "dumber"?

No! Please do not use the words like “stupid”, “dumb”, “brain rot”, "harm", "damage", "passivity", "trimming" and so on. It does a huge disservice to this work, as we did not use this vocabulary in the paper, especially if you are a journalist reporting on it.

- a preprint, not yet peer reviewed
- 54 people, one model, one 20-minute task
- EEG connectivity does not show harm

The authors' FAQ, brainonllm.com/faq; Kosmyna et al., Limitations and Future Work, p. 151

Skills: what do people still do themselves, and when do they practice?

Skills: what do people still do themselves, and when do they practice?

Judgment: can people tell when a task is outside the frontier?

Skills: what do people still do themselves, and when do they practice?

Judgment: can people tell when a task is outside the frontier?

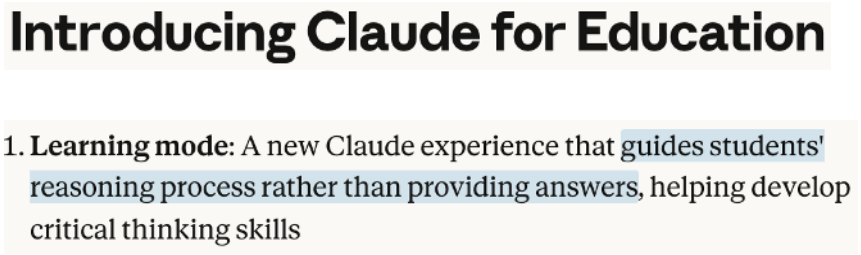
Ownership: whose ideas end up in the result?

Doing the task before using AI

“these findings support an educational model that delays AI integration until learners have engaged in sufficient self-driven cognitive effort”

Kosmyna et al., arXiv 2025, p. 151

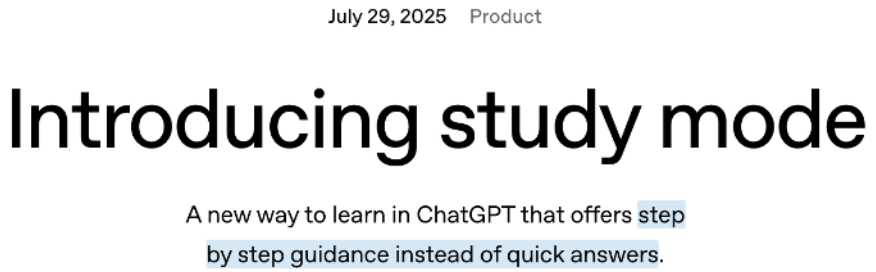
Learning mode and study mode



Introducing Claude for Education

1. **Learning mode**: A new Claude experience that guides students' reasoning process rather than providing answers, helping develop critical thinking skills

Anthropic, “Introducing Claude for Education,” 2025



July 29, 2025 Product

Introducing study mode

A new way to learn in ChatGPT that offers step by step guidance instead of quick answers.

OpenAI, “Introducing study mode,” 2025

Would you turn it on the night before a deadline?

Milestone 4: Decide what it does on its own

Due Tuesday of week 10

February 2026: METR's second study

Unfortunately, given participant feedback and surveys, we believe that the data from our new experiment gives us an unreliable signal of the current productivity effect of AI tools. The primary reason is that we have observed a significant increase in developers choosing not to participate in the study because they do not wish to work without AI, which likely biases downwards our estimate of AI-assisted speedup. We additionally believe there have been selection effects due to a lower pay rate (we reduced the pay from $150/hr to $50/hr), and that our measurements of time-spent on each task are unreliable for the fraction of developers who use multiple AI agents concurrently.

"I'm torn. I'd like to help provide updated data on this question but also I really like using AI!" — a developer from the original study early-2025 when asked to participate in the late-2025 study.

METR, "We are Changing our Developer Productivity Experiment Design," 2026

Takeaways

1. People can misjudge how AI changes their own **speed**.
2. Automating easy tasks takes away **practice** for hard ones.
3. AI helps on some tasks and **hurts** on others that look as hard.
4. Besides speed, check **skills**, **judgment** and **ownership**.

Next: Lecture 18, AI and Society

Read before class:

- Weidinger et al., “Taxonomy of Risks posed by Language Models,” FAccT 2022
- Samuelson, “Legal Challenges to Generative AI, Part II,” CACM 2023