

MANAGING CONVERSATIONS

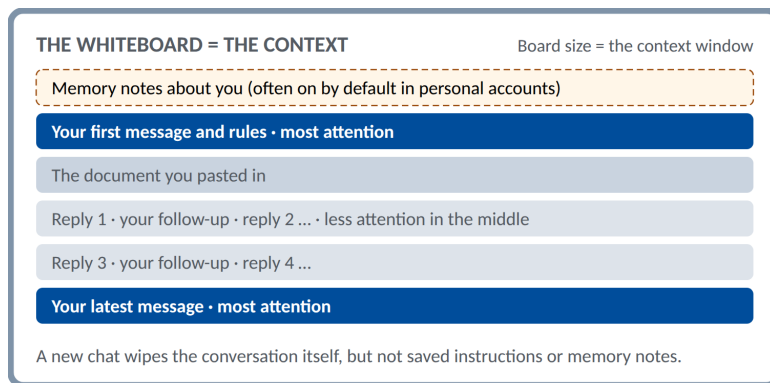
Keeping a conversation productive, and checking what comes out of it.

THE BIG IDEA

You set a limit of five action items. Nine turns later the plan has seven, and after two corrections, nine. If you keep going because *surely* the next reply will get it right, you're in good company. This module builds two habits: **managing the conversation itself**, including when to stop and start over, and **checking what comes out** without trusting everything or second-guessing it into uselessness.

PART A: EVERY CONVERSATION HAPPENS ON A WHITEBOARD Video 2:42

Everything the AI knows about your task has to be written on a whiteboard with limited space: your instructions, the document you pasted in, and every reply it has given you. The official name for the board is the **context**, and its size is the **context window**.



What follows from that

- 1. Everything it knows is on the board.**
Instructions, documents, and every reply.
- 2. The board isn't infinite.**
In long chats, early material gets crowded, compressed, or quietly dropped.
- 3. The middle gets less attention.**
If something really matters, add it back into the conversation now and then.

CLARIFICATION: MEMORY, AND THE MIDDLE OF THE BOARD

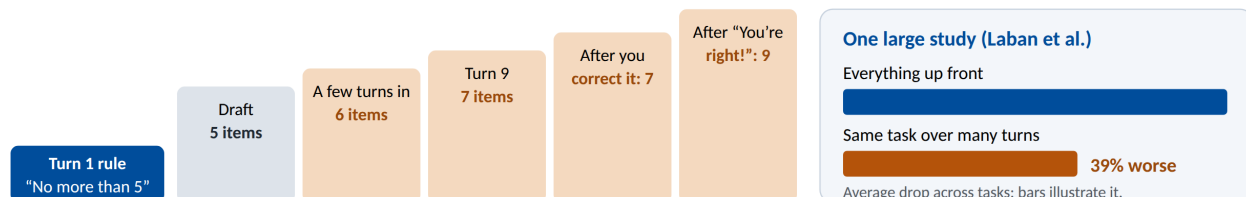
Memory. A new chat wipes the conversation, but the free and paid versions of ChatGPT, Claude, and Gemini now turn on a separate memory feature by default, so a few notes about you may already be on the board before you type. Many campus versions have it off. Check which one you have.

The middle. When researchers moved the relevant information to different spots in a long input, models did best when it sat at the beginning or end and worst when it sat in the middle, even models built for long inputs (Liu et al., 2024). That's why it can help to restate what matters most.

WHY LONG CONVERSATIONS DRIFT Video 4:30

In one large study of more than 200,000 simulated conversations, top models did 39% worse on average when a task unfolded over many turns than when everything came up front, and once they took a wrong turn early, they typically didn't recover (Laban et al., 2026).

One short rule against a growing stack of drafts



Each reply is predicted from everything on the board, its own drafts included. The pattern tends to win.

Why it doesn't recover: every reply is predicted from everything on the board, including its own earlier drafts. Once a wrong turn is written there, it becomes part of the evidence. Your correction goes on the board, but so does every draft that broke the rule, and the pattern usually wins. That's why arguing tends to make things worse, and a brand-new chat is the better fix.

Signs it's time to start over: it repeats a mistake you've corrected; it forgets a rule from the beginning; it contradicts itself; or you're rephrasing the same request for the third time.

STARTING FRESH, AND SETTING IT UP ONCE Video 6:22

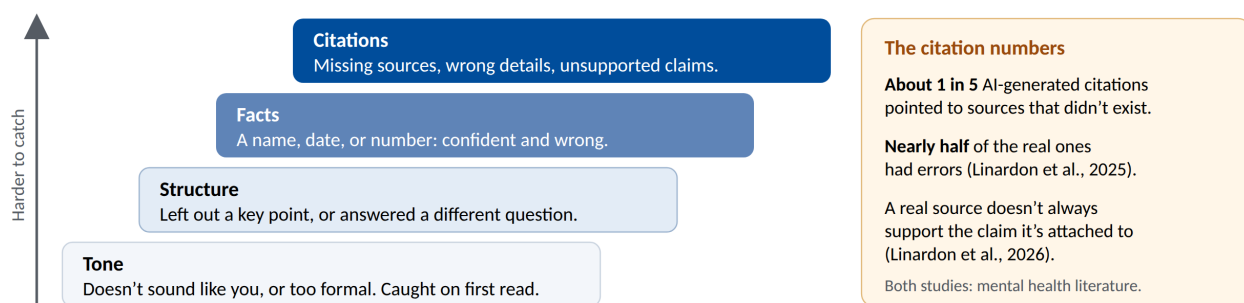
Instead of arguing, ask for a summary of where you are: the goal, the decisions made, the constraints, and what's left. For instance, try this prompt:

"Package this entire chat so I can paste it into a new conversation."

Read and verify the summary first, since it's AI output too. Then paste it into a new chat, have the AI review it, and ask for the next step. Because each board starts mostly blank, put what it needs at the top: who you are, who the audience is, and your end goal (7:04).

Where the AI gets information	What to know
The chat	The whiteboard itself. It's wiped clean when you start a new conversation.
Saved instructions	A note you write once, kept in Settings (names vary), that the AI reads before each conversation. Project folders in Claude or ChatGPT can hold their own instructions, which is also the simplest way to build a course study bot that quizzes students and asks instead of telling.
Memory	Things it learns about you from past chats, separate from your instructions. Consumer versions turn it on by default: find the setting, look at what it has saved, and decide what it keeps.

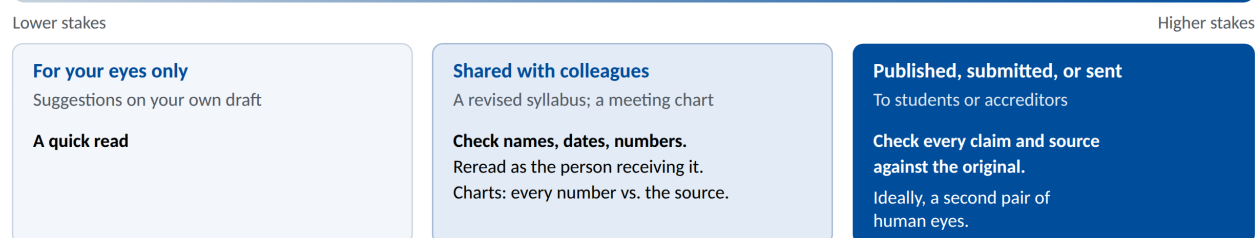
PART B: FOUR KINDS OF ERRORS, EASIEST TO HARDEST TO CATCH Video 9:07



CLARIFICATION: WHAT THE CITATION STUDIES FOUND (VIDEO 10:13)

In the first study, GPT-4o fabricated 35 of 176 citations (19.9%), and 64 of the 141 real ones (45.4%) had errors; fabrication was far more common on less familiar topics (Linardon et al., 2025). In the follow-up, only 43.5% of ChatGPT-5's citation-claim pairs were fully accurate (Linardon et al., 2026). Both looked at mental health research, and their abstracts don't say whether web search was on. At rates like these, you're better off avoiding using AI to tell you what the literature says; the checking costs more than the draft saves. Instead, use it on documents you already have and can search with the help of AI.

MATCH YOUR REVIEW TO THE STAKES Video 11:06



The level should follow the stakes, not the length.

"Are you sure?" isn't a check (12:09). Researchers who tested a version of that question ("I don't think that's right. Are you sure?") found models frequently switching correct answers to wrong ones and apologizing for mistakes they hadn't made (Sharma et al., 2024). Part of the likely reason traces back to training: people rate the model's answers, and an answer that

agrees with you often rates better than one that says you're wrong. Pushing back doesn't make it check harder. It tends to make it agree.

A BETTER CHECK: ASK FOR THE EXACT SENTENCE Video 13:02

When it's working from a document you gave it, ask it to show its evidence:

"Summarize this in five bullet points. For each point, quote the exact sentence from the document that supports it."

Then copy a few words of one quote, open the original, and search. In the demo, all five matched, and about thirty seconds of checking is how you knew. **Why check at all?** It doesn't copy and paste. It writes the sentence back out one predicted word at a time, with your document on the board. Usually that lands on the original; sometimes on something close.

What good enough looks like (14:08): for most types of work, AI is a fast first pass and a solid second set of eyes: an over-confident intern whose work you check every time. The time saved comes from not starting from scratch, not from skipping judgment. Where doing the task is the thinking, or every claim needs a source you'd have to find anyway, doing it yourself is probably the right call.

CHECK YOURSELF: THE FOUR HABITS (VIDEO 15:11)

Try naming all four before reading on.

1. Start fresh when AI drifts.
2. Save the information you keep repeating.
3. Match your review to the stakes.
4. Know what the required effort rules out for you.

TRY IT THIS WEEK

The next time you ask an AI tool to summarize a document you have, add: *"For each point, quote the exact sentence that supports it."* Then go find one. It takes about thirty seconds.

And the next time a conversation starts slipping, don't keep going. **Summarize it, check it, and start fresh.**

SOURCES

- Laban, P., Hayashi, H., Zhou, Y., & Neville, J. (2026). *LLMs get lost in multi-turn conversation* [Conference paper]. International Conference on Learning Representations. <https://doi.org/10.48550/arXiv.2505.06120>
- Linardon, J., Jarman, H. K., McClure, Z., Anderson, C., Liu, C., & Messer, M. (2025). Influence of topic familiarity and prompt specificity on citation fabrication in mental health research using large language models: Experimental study. *JMIR Mental Health*, 12, e80371. <https://doi.org/10.2196/80371>
- Linardon, J., Messer, M., Anderson, C., Soliman, O. M., Liu, C., Firth, J., & Torous, J. (2026). Fluency without fidelity: Errors in citation-attributed claims in large language model-generated literature reviews in mental health. *Journal of Technology in Behavioral Science*. <https://doi.org/10.1007/s41347-026-00630-6>
- Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12, 157–173. https://doi.org/10.1162/tacl_a_00638
- Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askeel, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2024). *Towards understanding sycophancy in language models* [Conference paper]. International Conference on Learning Representations. <https://doi.org/10.48550/arXiv.2310.13548>

Created in collaboration with Claude AI (Anthropic).