Why AI loses the plot in long chats

Six hacks I learnt the hard way.



They keep answers on track and stop you burning through your limits.

What a context window is

It's AI's working memory for one chat. It's measured in tokens, and a token is about ¾ of a word. Copilot, ChatGPT and Claude don't remember a chat the way we do. Every time you send a message, they re-read the whole chat from the top.

Why long chats go wrong

The more there is to re-read, the more AI misses. Early instructions get lost in the middle, and answers get vaguer or confidently wrong. Long chats also use up your limits faster, because every message carries the whole history.

Signs it's time for a new chat

- [ ] It forgets rules you gave it earlier
- [ ] It repeats itself or goes in circles
- [ ] It contradicts what it said before
- [ ] Answers sound sure but are wrong
- [ ] Replies slow right down

COPY THIS: HANDOFF PROMPT

I'm moving to a new chat. Write a handoff note I can paste in. Include the goal, what we've decided, what's left to do, and any rules or preferences I've given you. Keep it short.

SIX HACKS

1 One task, one chat

Start a new chat when the task or topic changes.

2 Hand off before you switch

Use the prompt above, then paste the note into a fresh chat.

3 Edit, don't pile on

Edit your original message instead of replying with fixes. The bad answer drops out.

4 Only add what it needs

Paste the section you're working on, not the whole document.

5 Ask if it's OK to carry on

On a roll? Ask AI how its context window is looking. It will tell you if you can keep going or should hand off.

6 Listen in voice mode

Using voice? You'll hear it when AI starts to struggle.