

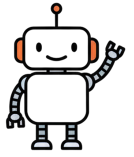


Why long chats forget the start

A one-page checklist, in plain English

WHY IT MATTERS

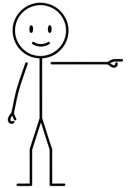
An AI can only look at so much text at once. In a very long chat, early messages may fall out of view. Here is how to work with that.



1 Know the term

The context window is all the text a model can look at when it writes a reply, including the reply itself. One provider's docs call it the model's "working memory".

Try: Think of it as the AI's desk.



2 Desk sizes differ

Different models have different context window sizes, and the numbers change over time. Look up your model in its official docs.

We do not quote numbers here on purpose.



3 Full desk, old stuff drops

Some chat apps manage the window "first in, first out", so the oldest messages can go first. That is why a rule from message one may stop working.

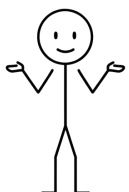
It depends on the app. Not every app does this.



4 Bigger is not magic

One provider's docs say more context is not automatically better: as the amount of text grows, accuracy and recall can get worse. Keep what you paste focused.

Try: Paste only the pages that matter.

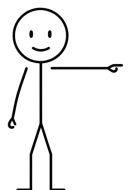


5 Repeat the key rule

Every so often, say your important rule again, or start a fresh chat and paste a short summary.

Try: "Reminder: call me Sam, and keep answers short."

Our suggestion, general good practice, not a tested rule.



Rule of thumb: if it matters, say it again.