

Why does AI forget what you just said? — Memory limits

Five-Minute AI Lessons · Sep 13, 2026

If it matters, say it again

In a nutshell

- 1 AI doesn't store a chat like a brain. Every time, it rereads **the whole conversation so far** from the top, then answers.
- 2 But **how much it can read at once (context length)** has a limit, so in a long chat the early part gets pushed out.
- 3 So for anything really important, **bringing it up again later in the chat** helps the AI and you.

In a long chat, old messages fall out of the window



With only 4 slots, what stays?

4 slots

Short chat

- ✓ Name: Sam
- ✓ Favorite color: blue
- ✗ Yesterday's weather
- ✗ A quick joke

Key info stays, the rest gets pushed out

4 slots

Medium-length chat

- ✓ Name: Sam
- ✓ Today's to-do list
- ✗ Favorite color: blue
- ✗ Yesterday's weather

The longer the chat, the oldest go first

4 slots

What if you say it again?

- ✓ Name: Sam (said again)
- ✓ Color: blue (said again)
- ✓ Today's to-do list
- ✗ A quick joke

Say key info again and it's back inside

As a chat gets longer, the oldest messages fall out of the window first.

※ Really, it's counted in much smaller units, not slots.

Just three questions

Q1

Why does AI 'forget' the start of a chat?

- ① Its feelings got hurt
- ② It can only read so much at once (its context)
- ③ It ignores you on purpose
- ④ It's testing you

Q2

(True / False) Even in a really long chat, AI always remembers exactly what you said first.

Q3

(True / False) If you repeat important info partway through a chat, AI uses it better.

Know the why, too

Q1

②

AI isn't a memory bank. It rereads the whole chat every time, so once the chat is longer than it can read, the start gets pushed out.

Q2

F

In a long chat, the early part can get pushed 'out of the window' and stop counting.

Q3

T

Anything you repeat is back 'inside the window', so the AI can use it more.

**If it matters,
say it again.**



Tomorrow: Prompts — why changing one question changes the whole answer