Mozout

Generated on 21/10/2026 16:01

KNOWLEDGE BASE · PUBLIC DOCUMENT

Why the AI forgets what you said earlier: context and long conversations

AI & Agents

CODE	VERSION	COUNTRY · LANGUAGE
RLX-INT-KB-4890	v2026.10.10	International · English

MozOut International
RLX Holdings (Pty) Ltd. · Group RLX Holdings
Tax no. (SARS) 943 907 6234 · Registration no. (CIPC) 2025/667953/07

https://www.my.mozout.com/knowledgebase/4890

Why the AI forgets what you said earlier: context and long conversations

AI & Agents Created on **October 10, 2026** 3 min read



AI “forgets” because it **has no memory of its own between messages**: each time it answers, it re-reads what has been sent in that conversation, up to a limit of text (the “context window”). When the conversation passes that limit, or when there is so much that what matters gets lost in the middle, the model no longer has the beginning in view. It is not carelessness, it is how it works.

The three most common reasons

What happens	Why	What to do
It forgets what you said at the start of a very long conversation	The context window filled and the tool cut off or summarised the beginning	Start a new conversation with a summary of your own of the decisions and important facts.
It contradicts an instruction you gave	You gave the instruction long ago, among a lot of text; or another instruction weighs more	Repeat it in the current request, short and at the end.
It does not remember yesterday’s conversation	Each conversation is separate unless the tool has a memory feature	Paste what matters, or use the tool’s memory and check what it stores.
It answers beside the point on a large document	The document did not fit whole, or the relevant part sat in the middle	Split it and ask in parts; point to the section you care about.

How to work so it does not get lost

1. **One conversation per subject.** When you change topic, open another. Dragging the history along confuses and costs more.

2. **Put the essentials at the start of the request** and repeat the rules that must not fail at the end.

3. **Ask for a summary halfway,** check it, and use it to start a new conversation: “These are the decisions already made: …”.

4. **Give the document and the question together,** and ask it to quote the passage it relied on. If it cannot quote, be suspicious.

5. **Use the tool’s standing instructions,** if it has them (a place for rules that apply to every conversation), instead of repeating them each time.

When used through an API, the history is **sent again with every request** and counts as text read, so a long conversation gets dearer and slower. See **how tokens are counted**.

There is also a common misunderstanding: thinking the AI “learns” from you during the conversation. It adapts to what is in the text of the conversation, but it does not know more once you close it, unless the tool says so explicitly and gives you control over what it keeps.

“Memory” features store data about you.
See what the tool remembers about you, where you can see it and how to delete it, before telling it about customers. See **the data terms**.

A bigger context window does not solve everything: when the text is very long, the model tends to pay less attention to the
middle
. A short, well-chosen summary often works better than dumping everything in.

RECOMMENDED

Want an assistant that works with your documents? See first how RAG works.

See RAG explained →

SEE ALSO

- **Tokens explained: how AI model APIs are counted** →
- **AI that answers from your own documents: RAG explained** →

Writing prompts for business tasks: replies, descriptions and summaries

Mozout

MozOut International

RLX Holdings (Pty) Ltd. · Group RLX Holdings

TAX NO. (SARS)	943 907 6234
REGISTRATION NO. (CIPC)	2025/667953/07
REGISTERED OFFICE	1111 Olive Street Saint Louis 63101, Missouri, United States
CONTACT	info@mozout.com · +27 78 823 3325 · www.mozout.com
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

Online version of this document

RLX-INT-KB-4890 · v2026.10.10 · Generated on 21/10/2026 16:01