

15. Verbatim Tail

Keeping the recent rounds whole

An AI agent compacts a window of four rounds but keeps the two most recent verbatim. The two older rounds hold 600 tokens between them, and the summary keeps half. Mark the window as it stands after the compaction, S: system, U: question, O: answer, C: summary.

Four rounds, the older two about to be summarized

1 - 1,000	S	S	(U1	O1	O1)	(U2	O2	O2)	(U3	O3
1,001 - 2,000	O3)	(U4	O4	O4)						

Step 1: Take the share the summary keeps

$$0.5 \times 600 = 300 \text{ tokens, or 3 cells}$$

Step 2: Mark the window, summary in place of the older rounds

1 - 1,000	S	S	C	C	C	(U3	O3	O3)	(U4	O4
1,001 - 2,000	O4)									

15. Verbatim Tail (continued)

Practice 1

An AI agent compacts a window of four rounds but keeps the two most recent verbatim. The two older rounds hold 800 tokens between them, and the summary keeps a quarter. Mark the window as it stands after the compaction, S: system, U: question, O: answer, C: summary.

Four rounds, the older two about to be summarized

1 - 1,000	S	S	S	(U1	O1	O1	O1)	(U2	O2	O2
1,001 - 2,000	O2)	(U3	O3	O3	O3)	(U4	O4	O4	O4)	

Step 1: Take the share the summary keeps

Step 2: Mark the window, summary in place of the older rounds

1 - 1,000										
1,001 - 2,000										

Practice 2

An AI agent compacts a window of five rounds but keeps only the most recent verbatim. The four older rounds hold 1,200 tokens between them, and the summary keeps a quarter. Mark the window as it stands after the compaction, S: system, U: question, O: answer, C: summary.

Five rounds, the older four about to be summarized

1 - 1,000	S	S	(U1	O1	O1)	(U2	O2	O2)	(U3	O3
1,001 - 2,000	O3)	(U4	O4	O4)	(U5	O5	O5)			

Step 1: Take the share the summary keeps

Step 2: Mark the window, summary in place of the older rounds

1 - 1,000										
1,001 - 2,000										