

10. Truncation

Crossing out the oldest turns

An AI agent has a 4,000-token context window holding a 300-token system prompt and 7 turns of 500 tokens each. A new 1,500-token message arrives. How many of the earliest turns must be crossed out to make room?

Step 1: How much room is short

$1,500 - 200 = 1,300 \text{ tokens short}$

Step 2: Turns to drop

$1,300/500 = 2.6, \text{ so } 3 \text{ turns}$

Step 3: Cross out the earliest turns with a line

1 - 1,000	S	S	S	(	U	U	O	O	O	)	X	U	U
1,001 - 2,000	O	O	O	X	U	U	O	O	O	X	U	U	
2,001 - 3,000	O	O	O	X	U	U	O	O	O	X	U	U	
3,001 - 4,000	O	O	O	X	U	U	O	O	O	)			

10. Truncation (continued)

Practice 1

An AI agent has a 4,000-token context window holding a 200-token system prompt and 9 turns of 400 tokens each. A new 1,000-token message arrives. How many of the earliest turns must be crossed out to make room?

Step 1: How much room is short

Step 2: Turns to drop

Step 3: Cross out the earliest turns with a line

1 - 1,000	S	S	(	U	O	O	O	)	U	O	O	O	
1,001 - 2,000	(	U	O	O	O	)	U	O	O	O	)	U	O
2,001 - 3,000	O	O	)	U	O	O	O	)	U	O	O	O	
3,001 - 4,000	(	U	O	O	O	)	U	O	O	O	)		

Practice 2

An AI agent has a 3,000-token context window holding a 300-token system prompt and 8 turns of 300 tokens each. A new 1,400-token message arrives. How many of the earliest turns must be crossed out to make room?

Step 1: How much room is short

Step 2: Turns to drop

Step 3: Cross out the earliest turns with a line

1 - 1,000	S	S	S	(	U	O	O	)	U	O	O	)	U
1,001 - 2,000	O	O	)	U	O	O	)	U	O	O	)	U	O
2,001 - 3,000	O	)	U	O	O	)	U	O	O	)			