

14. The Compaction Threshold

Picking the number from a round

An AI agent has a 2,000-token context window. Each round is estimated to include a 100-token question and a 300-token answer. Fill the window with rounds, and leave room for at least one full round. Set the trigger so compaction fires there automatically. S: system, U: question, O: answer.

Step 1: Mark rounds, leaving room for one more

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

Step 2: Take that position as a share of the window

$14 \div 20 = 0.7$

Step 3: Fill in the value in the config

```
context_management = {  
  "context_window": 2000,  
  "compaction_ratio": ,  
}
```

14. The Compaction Threshold (continued)

Practice 1

An AI agent has a 2,000-token context window. Each round is estimated to include a 100-token question and a 200-token answer. Fill the window with rounds, and leave room for at least one full round. Set the trigger so compaction fires there automatically. S: system, U: question, O: answer.

Step 1: Mark rounds, leaving room for one more

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

Step 2: Take that position as a share of the window

Step 3: Fill in the value in the config

```
context_management = {  
    "context_window": 2000,  
    "compaction_ratio": ,  
}
```

14. The Compaction Threshold (continued)

Practice 2

An AI agent has a 2,000-token context window. Each round is estimated to include a 200-token question and a 200-token answer. Fill the window with rounds, and leave room for at least one full round. Set the trigger so compaction fires there automatically. S: system, U: question, O: answer.

Step 1: Mark rounds, leaving room for one more

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

Step 2: Take that position as a share of the window

Step 3: Fill in the value in the config

```
context_management = {  
    "context_window": 2000,  
    "compaction_ratio": ,  
}
```