

5. Headroom

What is left of the window

An agent has a 4,000-token context window. The system prompt and conversation history already use 1,200 tokens. The document to read is 55 paragraphs, each about 50 tokens. Can the agent read the document in this call?

Step 1: Space available for the document

$$4,000 - 1,200 = 2,800 \text{ tokens}$$

Step 2: Estimate the document

$$55 \times 50 = 2,750 \text{ tokens}$$

Step 3: Does it fit?

$$2,750 < 2,800 \text{ — yes, but barely.}$$

Practice 1

An agent has a 4,000-token context window. The system prompt and conversation history already use 600 tokens. The document to read is 8 pages, each about 500 tokens. Can the agent read the document in this call?

Step 1: Space available for the document

Step 2: Estimate the document

Step 3: Does it fit?

5. Headroom (continued)

Practice 2

An agent has a 4,000-token context window. The system prompt and conversation history already use 1,000 tokens. The document to read is 18 slides, each about 150 tokens. Can the agent read the document in this call?

Step 1: Space available for the document

Step 2: Estimate the document

Step 3: Does it fit?