

6. The Price of a Call

An AI agent processes a call of 600 tokens. The model is priced at 2 cents per 100 tokens. What does this call cost?

Step 1: Count the 100-token blocks

$$600/100 = 6 \text{ blocks}$$

Step 2: Multiply by the price per block

$$6 \times 2 \text{ cents} = 12 \text{ cents}$$

Practice 1

An AI agent processes a call of 2,000 tokens. The model is priced at 3 cents per 100 tokens. What does this call cost?

Step 1: Count the 100-token blocks

Step 2: Multiply by the price per block

Practice 2

An AI agent processes a call of 800 tokens. The model is priced at 5 cents per 100 tokens. What does this call cost?

Step 1: Count the 100-token blocks

Step 2: Multiply by the price per block

7. Two Kinds of Token

An AI agent is sent the input “why do cats purr?” and returns the output “a cat purrs when it feels happy!”. Tokenize the input and the output, then count the tokens of each kind.

Step 1: Tokenize the input

why	do	cats	purr	?	
-----	----	------	------	---	--

Step 2: Tokenize the output

a	cat	purrs	when	it	feels
happy	!				

Step 3: Count the tokens of each kind

input = 5, output = 8

Practice 1

An AI agent is sent the input “how many moons does the planet Jupiter have?” and returns the output “the planet Jupiter has ninety five moons!”. Tokenize the input and the output, then count the tokens of each kind.

Step 1: Tokenize the input

Step 2: Tokenize the output

Step 3: Count the tokens of each kind

7. Two Kinds of Token (continued)

Practice 2

An AI agent is sent the input “is the sun a star?” and returns the output “yes, the sun is a star.”. Tokenize the input and the output, then count the tokens of each kind.

Step 1: Tokenize the input

--	--	--	--	--	--

Step 2: Tokenize the output

Step 3: Count the tokens of each kind

--

8. Input or Output?

A model receives the question “what is the plan” and replies “first gather data then analyze results”. Label each token I for input or O for output.

Step 1: Label each token I or O

I	I	I	I	O	O	O	O	O
what	is	the	plan	first	gather	data	then	analyze

O

results

Step 2: Count the tokens of each kind

input = 4 tokens · output = 6 tokens

Practice 1

A model receives the task “summarize the meeting notes” and replies “the team agreed to ship by Friday”. Label each token I for input or O for output.

Step 1: Label each token I or O

summarize	the	meeting	notes	the	team	agreed
-----------	-----	---------	-------	-----	------	--------

to	ship	by	Friday
----	------	----	--------

Step 2: Count the tokens of each kind

8. Input or Output? (continued)

Practice 2

A model receives the request “list the tasks assigned to Alice” and replies “Alice has three open tasks”. Label each token I for input or O for output.

Step 1: Label each token I or O

list	the	tasks	assigned	to	Alice	Alice
------	-----	-------	----------	----	-------	-------

has	three	open	tasks
-----	-------	------	-------

Step 2: Count the tokens of each kind

9. The Two-Part Bill

An AI agent sends 600 input tokens and receives 200 output tokens. Input is priced at 2 cents per 100 tokens; output is priced at 8 cents per 100 tokens. What does the call cost in total?

Step 1: Cost of the input tokens

$$600/100 \times 2 = 12 \text{ cents}$$

Step 2: Cost of the output tokens

$$200/100 \times 8 = 16 \text{ cents}$$

Step 3: Add the two costs

$$12 + 16 = 28 \text{ cents}$$

Practice 1

An AI agent sends 1,000 input tokens and receives 500 output tokens. Input is priced at 2 cents per 100 tokens; output is priced at 8 cents per 100 tokens. What does the call cost in total?

Step 1: Cost of the input tokens

Step 2: Cost of the output tokens

Step 3: Add the two costs

9. The Two-Part Bill (continued)

Practice 2

An AI agent sends 400 input tokens and receives 300 output tokens. Input is priced at 2 cents per 100 tokens; output is priced at 8 cents per 100 tokens. What does the call cost in total?

Step 1: Cost of the input tokens

Step 2: Cost of the output tokens

Step 3: Add the two costs

10. End-to-End Call Cost

An AI agent receives a 200-word prompt and returns 200 output tokens. Input tokens are estimated at 1.5 per word. Input is priced at 2 cents per 100 tokens and output at 8 cents per 100 tokens. What does the call cost in total?

Step 1: Estimate the input tokens

$$200 \times 1.5 = 300 \text{ tokens}$$

Step 2: Cost of the input tokens

$$300/100 \times 2 = 6 \text{ cents}$$

Step 3: Cost of the output tokens

$$200/100 \times 8 = 16 \text{ cents}$$

Step 4: Add the two costs

$$6 + 16 = 22 \text{ cents}$$

10. End-to-End Call Cost (continued)

Practice 1

An AI agent receives a 100-word prompt and returns 600 output tokens. Input tokens are estimated at 2.0 per word. Input is priced at 2 cents per 100 tokens and output at 8 cents per 100 tokens. What does the call cost in total?

Step 1: Estimate the input tokens

Step 2: Cost of the input tokens

Step 3: Cost of the output tokens

Step 4: Add the two costs

Practice 2

An AI agent receives a 200-word prompt and returns 100 output tokens. Input tokens are estimated at 2.5 per word. Input is priced at 2 cents per 100 tokens and output at 8 cents per 100 tokens. What does the call cost in total?

Step 1: Estimate the input tokens

Step 2: Cost of the input tokens

Step 3: Cost of the output tokens

Step 4: Add the two costs