

GLM-5.3

Zhipu AI

AT A GLANCE

PRICE PER 1M

\$0.05 / \$7

input / output · cached
\$0.045

CONTEXT

1.05M

tokens in

MAX OUTPUT

944K

tokens per reply

TAKES

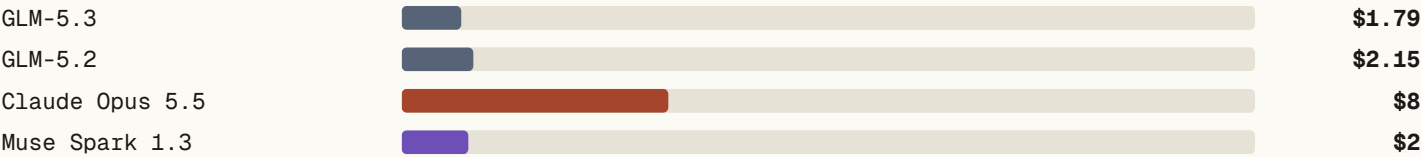
text

returns text

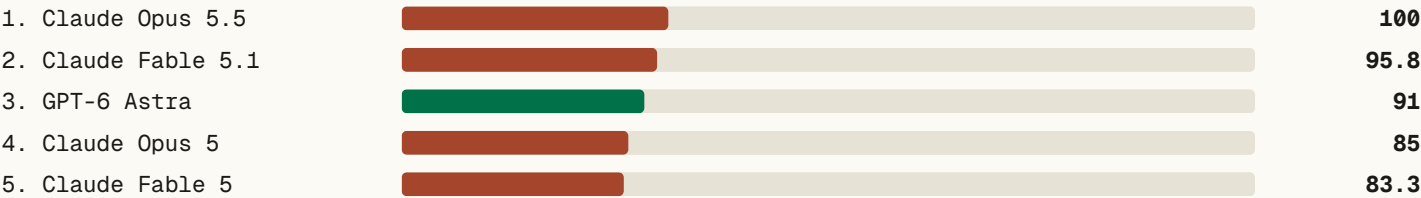
HOW IT COMPARES

MODEL	ROLE	IN / 1M	OUT / 1M	CONTEXT	BENCH INDEX
GLM-5.3	new	\$0.05	\$7	1.05M	47.8
GLM-5.2	it follows	\$1.40	\$4.400	1M	23.3
Claude Opus 5.5	today's #1	\$4	\$20	1M	100
Muse Spark 1.3	best value	\$1.25	\$4.25	1.05M	79.1
vs GLM-5.2	input 96% cheaper , output 59% more the model it follows: \$1.40 / \$4.400, bench index 23.3; context 1M				
vs Claude Opus 5.5	input 99% cheaper , output 65% cheaper today's #1: \$4 / \$20, bench index 100; context 1M				
vs Muse Spark 1.3	input 96% cheaper , output 65% more best value: \$1.25 / \$4.25, bench index 79.1				

PRICE PER 1M TOKENS, BLENDED 3 IN : 1 OUT



WHERE THE PUBLIC LEADERBOARDS STAND TODAY



GLM-5.3 sits at 47.8 on the bench index. The bench index is the median of each model's percentile across the public leaderboards that rate it (LMArena, LiveBench, SimpleBench, LLM Stats, Vals AI, Artificial Analysis).

Sources: price, limits and modalities for GLM-5.3 as the OpenRouter catalog lists z-ai/glm-5.3; the other models as their makers list them (models.dev); rankings from the ArtificialWatch Bench Index. Facts only: no forecast of ours, no claims from the lab's announcement.