AI MODEL ROUTING CHEATSHEET

The marketer's guide to which model, for which job, at what cost. Prices per 1M tokens.

MODEL	PRICE IN / OUT	USE IT FOR
CLAUDE		Claude API pricing
Opus 5.5	$4 / $20	Planning and strategy
Sonnet 5.5	$2 / $10	Drafting and editing
Haiku 5.5 Prompts up to 100K tokens	$0.10 / $0.50	Tagging and checking against a list
CHATGPT		OpenAI API pricing
GPT-6.1 Terra Fast mode $2 / $10	$1 / $5	Pulling fields into a table
GPT-6 Luna	$0.10 / $0.50	Sorting with clear rules
GEMINI		Gemini API pricing
Gemini 3.1 Pro Preview	$2 / $12	Research and visual checks
Gemini 3.5 Flash-Lite	$0.30 / $2.50	Low-cost checking
OTHER PROVIDERS		Direct or via OpenRouter
Grok 4.7	$2 / $6	Research and short-copy hooks
MiniMax M3.1 Flash MiniMax direct API	$0.15 / $0.60	Bulk tagging at volume

UP TO 40X APART

Opus 5.5 output costs $20 per 1M tokens. Haiku 5.5 and GPT-6 Luna cost $0.50. Decision flow and scorecard on page 2.

Standard API text prices, USD per 1M tokens, checked Nov 5, 2026 on the Anthropic, OpenAI, Google, xAI and MiniMax pricing pages. Excludes caching and batch discounts.

© 2026 MarketingAlec · Skip the AI hype. Get AI results.

ROUTE IT, THEN TEST IT

Two questions pick the tier. Ten real examples tell you whether the cheaper model can keep the job.

THE TWO-QUESTION ROUTING TEST

QUESTION 1

Does the job need judgment?

"Decide what matters here."
Yes: Opus 5.5.
No: go to question 2.

→

QUESTION 2

Does it need your voice?

"Say this well."
Yes: Sonnet 5.5.
No: it's a checking job.

→

EVERYTHING ELSE

Send it down the table

"Check this against that list."
Haiku 5.5, GPT-6 Luna, Gemini Flash-Lite, MiniMax M3.1 Flash or cheaper.

Don't flip it: a classification model asked to "decide what matters" gives a confident, wrong answer.

BEFORE YOU SWITCH: RUN 10 REAL EXAMPLES THROUGH BOTH

Same brief in both models. Check each result against the source, not against the other model.

TRACK	USUAL MODEL	CHEAPER MODEL
Results that meet your rules Out of 10	___ / 10	___ / 10
Time to review and fix Average minutes per result	___ min	___ min
Retries needed Extra attempts across all 10	___	___
Cost of the finished job Under your tool's billing, retries included	$ ___	$ ___

Keep the cheaper model if it passes as often and the total cost, with retries and your fix time, comes out lower.
Prompts for the 3 jobs: marketingalec.com/resources/the-marketers-guide-to-ai-model-routing

© 2026 MarketingAlec · Skip the AI hype. Get AI results.