OPUS or SONNET?

Half the price per token is not half the bill.

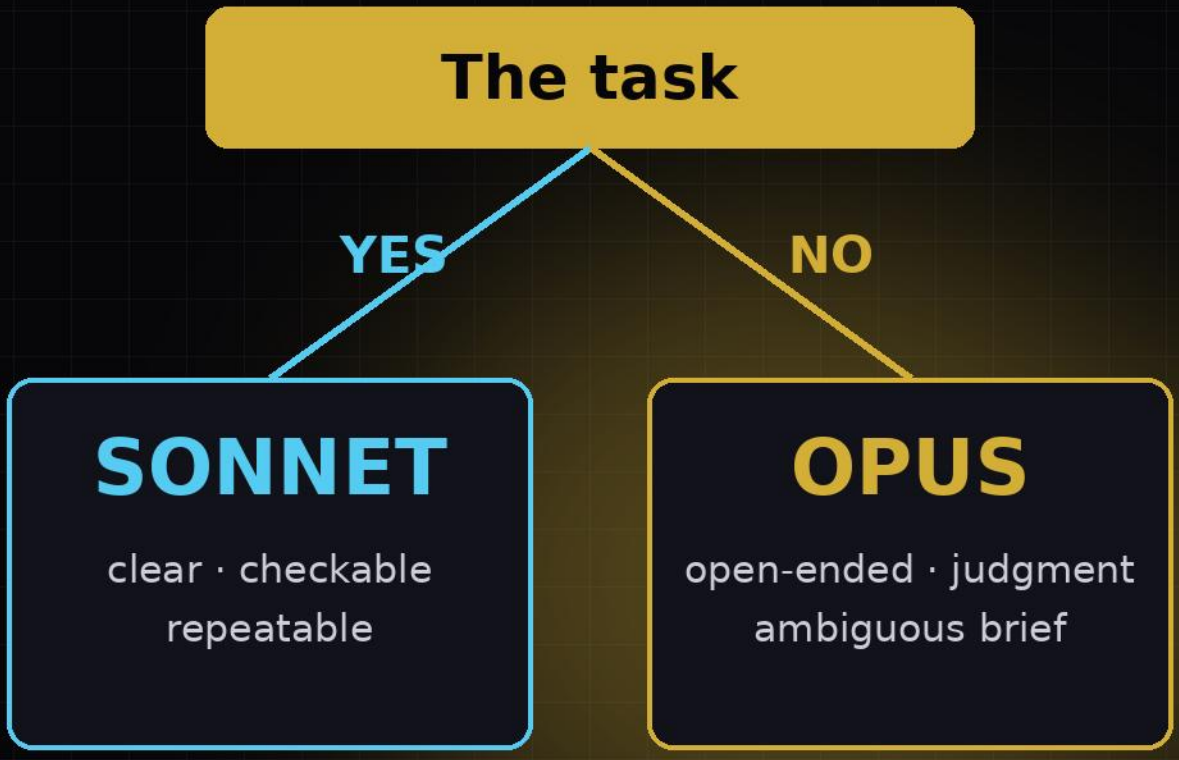
OPUS 5.5	SONNET 5.5
$4 / $20	$2 / $10

A rulebook for finance & accounting teams.

Price per 1M input / output tokens. Source: Anthropic, Sept 2026.

01 · THE ONE-LINE RULE

Can you write the definition of done?



If a reviewer can check it in two minutes, it's a Sonnet job.

02 · THE RATE CARD

What you pay per million tokens.

USD / 1M TOKENS	OPUS 5.5	SONNET 5.5
Input	$4.00	$2.00
Output	$20.00	$10.00
Cache write (5 min)	$5.00	$2.50
Cache read	$0.20	$0.20
Batch API	−50%	−50%
Context window	1M	1M

Input and output: half price. Cache: same.
Now watch the invoice →

Source: Anthropic model docs and launch pages, Sept 2026.

03 · HEAD TO HEAD

Opus 5.5 vs Sonnet 5.5.

	OPUS	SONNET
Knowledge work GDPval-AA	1846	1844
Long docs & deliverables AA-Briefcase	1822	1811
Chart reading Chartography	64.4%	61.6%
Computer use OSWorld 2.1	81.8%	80.1%
Hard reasoning HLE, tools	67.7%	64.5%
Coding sessions CursorBench 4	57.8%	55.5%
Terminal agents Terminal-Bench 4	66.4%	70.6%
Facts from memory AA-Omniscience accuracy	66%	54%
Guessing when unsure AA-Omniscience hallucination	59%	47%
Cost per task, max Artificial Analysis	$5.98	$7.60

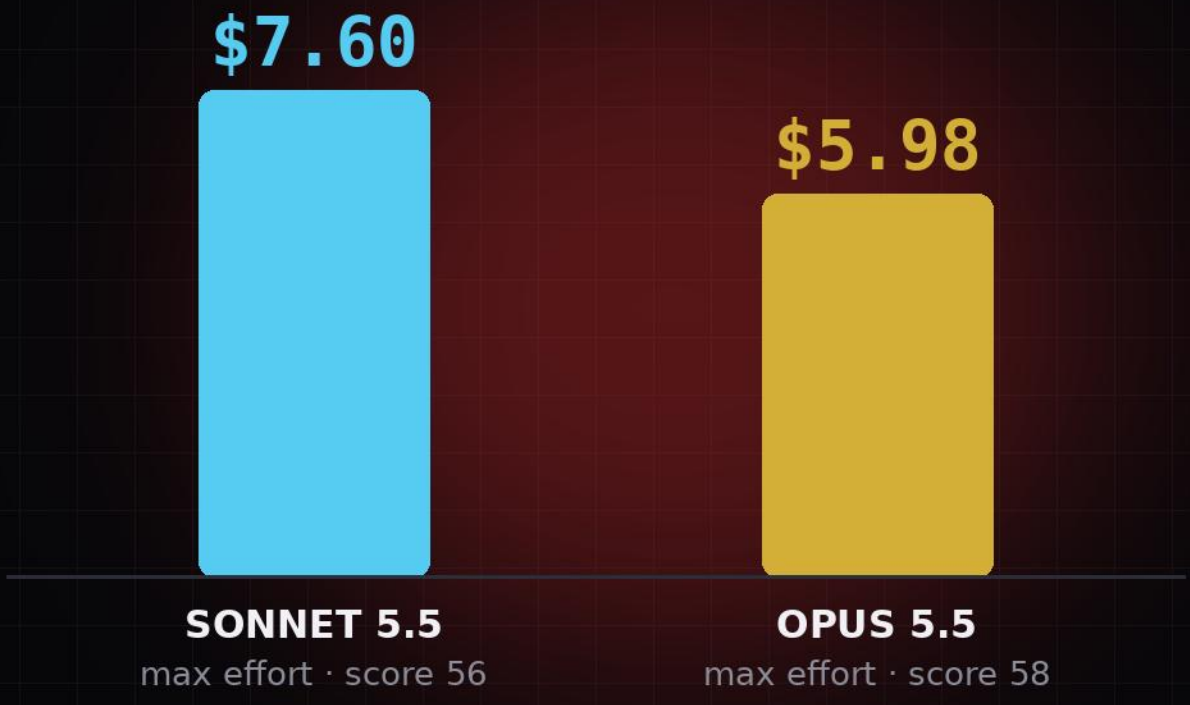
Opus leads 8 of 10, mostly by a hair.

The gaps that matter: facts and cost at max.

Benchmarks: Anthropic launch table (Terminal-Bench: Opus xhigh, Sonnet max). Facts & cost: Artificial Analysis. Haiku 5.5 not released yet.

04 · THE MAX-EFFORT TRAP

Push Sonnet to max and it costs more.



Cost per task, Artificial Analysis index.
Sonnet wrote ~60% more output tokens.

Cranking Sonnet to max? You needed Opus.

Source: Artificial Analysis Intelligence Index, Sept 2026. Your workload will differ.

05 · THE SONNET DESK

The month-end workhorse.

- ✓ Bank & intercompany reconciliations
- ✓ Accrual and prepayment schedules
- ✓ Budget-vs-actual variance tables
- ✓ Fixing Excel formulas and macros
- ✓ Invoice & contract data extraction
- ✓ Monthly deck from a fixed template

Anthropic test: a 10-slide operating review from earnings files, judged ready to send.

Task list: SKYTH's mapping of Anthropic's 'well-scoped everyday tasks'.

06 · THE OPUS DESK

Where judgment is the product.

- M&A / valuation model + IC memo
- Revenue-recognition judgments
- Cash-flow scenarios & strategy
- Audit findings: root cause
- Vague brief from the CFO
- Designing the close process itself

Early tester: Opus caught an error in their own eval instructions. No model had before.

Task list: SKYTH's mapping. Tester quote published by Anthropic.

07 · FACTS FROM MEMORY

Neither is your tax table.





- Tax rates, IFRS paragraphs, FX: from documents
- Every figure traced to a source line
- Numbers you sign: a human checks them

Source: Artificial Analysis AA-Omniscience, Sept 2026. Both models still get facts wrong.

08 · THE ROUTING MATRIX

Stakes × clarity.

STAKES ↑	← CLEAR SPEC	OPEN-ENDED →
	CLEAR · HIGH STAKES **SONNET** `high effort` + human sign-off	OPEN · HIGH STAKES **OPUS** `xhigh effort` + human sign-off
	CLEAR · LOW STAKES **SONNET** `low / medium` spot checks	OPEN · LOW STAKES **OPUS** `medium effort` review the output

Opus designs the process. Sonnet runs it.

SKYTH framework. Effort levels: low · medium · high · xhigh · max.

09 · WHAT FINANCE TESTERS SAW

Early reports, vendor-published.

SONNET 5.5 — 2,441 finance tasks

Best quality-to-cost of the 7 models one tester ran for high-volume work.

SONNET 5.5 — financial services

Rechecks data in source documents, catching errors the last model missed.

OPUS 5.5 — end-to-end finance workflows

Covered 86.6% of an expert rubric. Best-ever citation recall for the tester.

Early-tester quotes from Anthropic launch pages; not independently verified.

THE VERDICT

Sonnet runs the close.
Opus makes the calls.
A human signs.

Default to Sonnet at medium effort.
Reaching for max? Switch to Opus.

Full rulebook with sources:
skyth.ai/library/opus-or-sonnet-claude-5-5-finance-rulebook

Link in the post ↑