

Which AI for Which Job

Every AI app you use (ChatGPT, Gemini, Claude and the rest) offers a faster, cheaper model and a bigger, pricier one. This card helps you pick the right one for each job, so you stop paying flagship prices for work that a smaller model handles just as well.

1 Run the "done" test first

Before you open any app, finish this sentence: **"This job is done when I have ____."** If you can fill the blank (a one-page summary, a formatted table, a reply to this email), the job is **well-scoped** and a smaller model will usually do it well. If you can't, you're asking the AI to help you **work out the goal**, and that is where the bigger model earns its price.

2 Match the job to a tier

THE JOB	TIER TO START WITH	THINKING OR EFFORT SETTING
Quick answers, rewording a sentence, a short reply	Fast	Lowest. Speed matters more than depth here.
Summarising a long document or meeting notes	Mid	Default. Raise it one step only if the summary misses key points.
Formatting a spreadsheet, slides, or a document to a template	Mid	Default. The template is your "done", so extra thinking rarely helps.
First draft of an email, post, or sales page from a clear brief	Mid	One step above default.
One decision that needs many files or sources weighed together	Flagship	Default. Its judgment is what you're paying for.
Brand voice, design direction, or anything that has to "feel right"	Flagship	Default, and give it examples of what you like.
Open-ended planning where you're not sure what you need yet	Flagship	Default. Ask it to propose the goal before the work.

3 Four rules that save you money

Start one tier lower than feels safe.

If the result is good, you've paid less. If it isn't, you've lost a minute, and you move up one tier.

Change one thing at a time.

Raise the effort setting or change the model, never both at once, so you learn which one fixed the result.

A better brief beats a bigger model.

Most weak answers come from vague requests. Add the "done" sentence and an example before you upgrade.

Never max out a cheap model.

Past a point it costs more per job than the flagship and still does slightly worse. If you reach for "max", switch up a tier (real numbers on page 2).

Your current models

Model names change every few months, so this page is the part that gets updated.

Checked 29 September 2026

A Claude, verified

TIER	MODEL	PRICE PER MILLION TOKENS (INPUT / OUTPUT)
Mid	Sonnet 5.5	\$2 / \$10
Flagship	Opus 5.5	\$4 / \$20

The trap, in real numbers: in Artificial Analysis's tests, Sonnet 5.5 cost about \$1.08 per task at its High effort setting but \$7.60 at Max, which is more than Opus 5.5's \$5.98 at Max, while scoring slightly lower (56 against 58). Effort settings run low, medium, high, xhigh and max, and the Claude apps default to medium.

Sources: anthropic.com/claude-sonnet-5-5 · platform.claude.com (Opus 5.5 overview, Sonnet 5.5 migration guide) · artificialanalysis.ai (Claude Sonnet 5.5 and Opus 5.5 pages). Prices are for the API; app plans are billed as usage limits rather than per token.

B Fill in the apps you use

App 1: _____
 Fast model: _____
 Mid model: _____
 Flagship: _____
 Where's the effort or "thinking" switch?

App 2: _____
 Fast model: _____
 Mid model: _____
 Flagship: _____
 Where's the effort or "thinking" switch?

Tip: open the model menu in each app. The smallest or "fast" option is your Fast tier, the default is usually Mid, and the top or "pro" option is your Flagship.

C Your one-week log

JOB	APP AND MODEL	SETTING	GOOD ENOUGH?