

There is a menu beside the message box. Does it matter which one I pick?

Two dials: which model, and how hard it thinks. Bigger is not always better.

thirdbrain.

Start →

1 Where the settings live

The menu beside the message box usually holds two separate controls: a model selector, and a setting for thinking or effort. The labels differ between Claude, ChatGPT and Gemini. If you have never changed them, every answer you have had has used whatever the product picked for you.

Microsoft Copilot works differently here. Instead of two separate controls it offers a single combined choice, so you pick one option rather than setting a model and an effort level separately.

2 Choose by the job, not the model name

Choose an AI model by the difficulty of the job, not by memorising the product name.

Model names and versions change quickly. The description beside the model is more useful than the label. Most model lists fall into three practical groups: fastest answers, everyday work and hardest problems.

Choose the smallest model that can do the job well.

3 Effort controls how long it works

Turn effort down and the model answers faster.

Turn it up and the model spends longer reasoning before it responds.

Higher effort also uses more of your allowance and may cost more in products that charge by usage. Some tools call it thinking, reasoning or deep think.

Same model. Different amount of thinking.

4 Where to start, and when to change

Start with the everyday model and its normal effort setting, then move down for simple work or up for difficult reasoning.

01 Start with the everyday model and its normal effort setting.

02 Move down for simple work where speed matters.
Summaries, first drafts, extraction and routine questions.

03 Move up for work that needs careful reasoning.
Strategy, complex analysis, competing constraints or a high-stakes recommendation.

04 Compare the result before assuming the bigger model is better.

One job, one chat, one model.

5 A practical model map

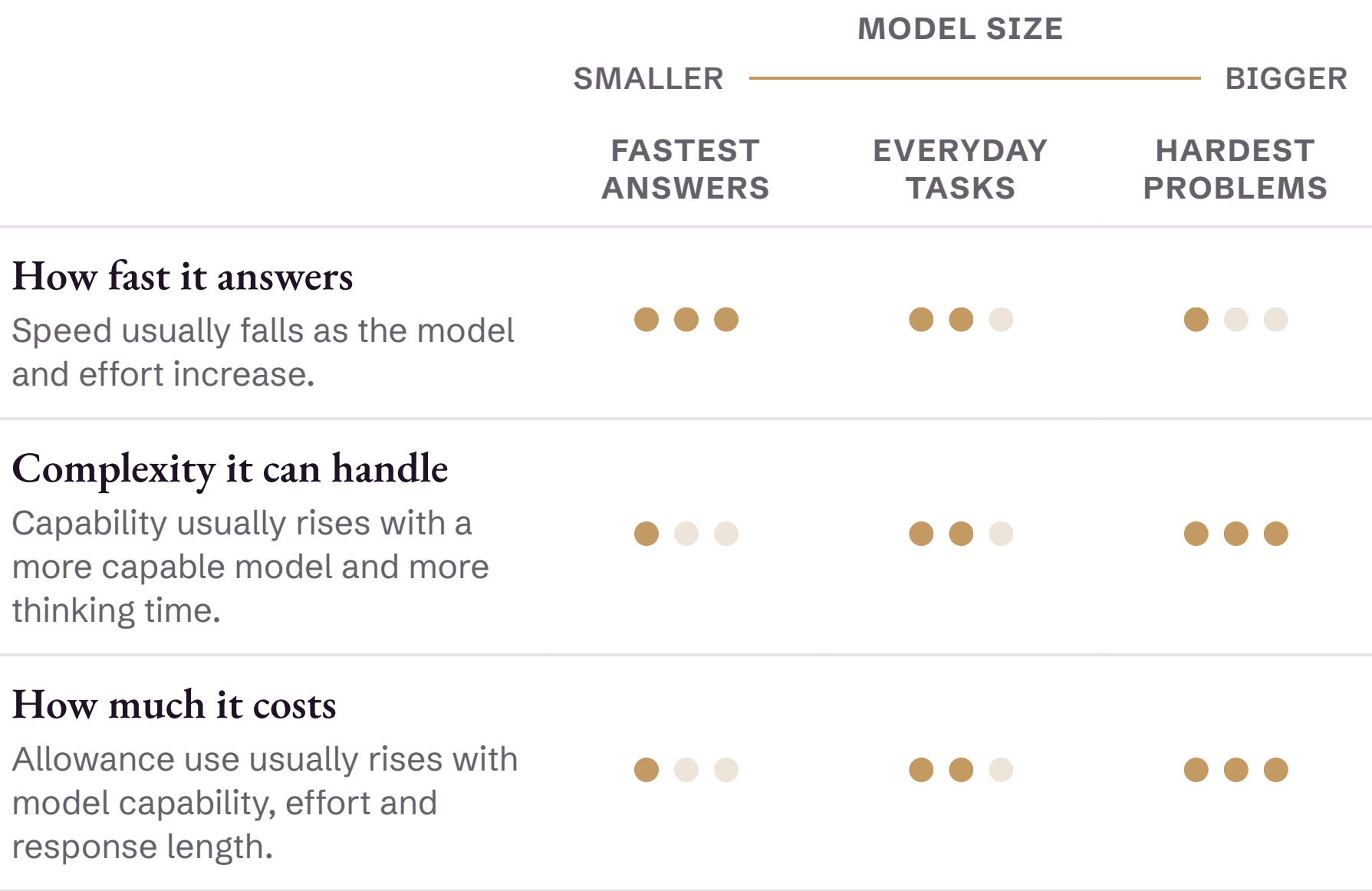
Examples current to 7 September 2026. Availability varies by product and plan.

Fastest answers	Everyday tasks	Hardest problems
Simple work where response time matters most.	The best starting point for most work.	Complex reasoning, analysis and multi-step work.
Claude Haiku 4.5	Claude Sonnet 5	Claude Opus 5 or Fable 5.1
GPT-5.6 Luna	GPT-5.6 Terra	GPT-5.6 Sol or GPT-6 Astra
Gemini 3 Flash-Lite	Gemini 3 Flash	Gemini 3 Pro

OpenAI launched GPT-6 Astra on 3 September 2026. It appears as GPT-6 Pro in ChatGPT. GPT Terra and Luna are available in Work, Codex and OpenAI's developer tools, not standard ChatGPT conversations. Model lists will keep changing.

6 What each model trades

As models get bigger, answers usually get slower and more expensive, and the work they can handle gets harder.



These are coarse groups, not scores. On a monthly chat plan, cost often shows up as more of your allowance rather than a separate bill.

7 Try this in three minutes

Run the same prompt on the fastest model and the most capable model, then decide whether the slower answer earned the extra time.

-
- 01 Open a new chat. Choose the fastest model and lowest effort available, then paste a real prompt.
 - 02 Open another new chat. Choose the most capable model and highest effort available, then paste the same prompt.
 - 03 Time both answers and read them side by side.
 - 04 Decide whether the slower answer earned the extra time and allowance.
-

What did you notice?

A FIRST CONVERSATION

Want help applying this to your work?

Bring one real task. We will work through it together and sharpen your setup as we go. No deck, no pitch.

thirdbrain.com.au

hello@thirdbrain.com.au

thirdbrain.

9 / 9