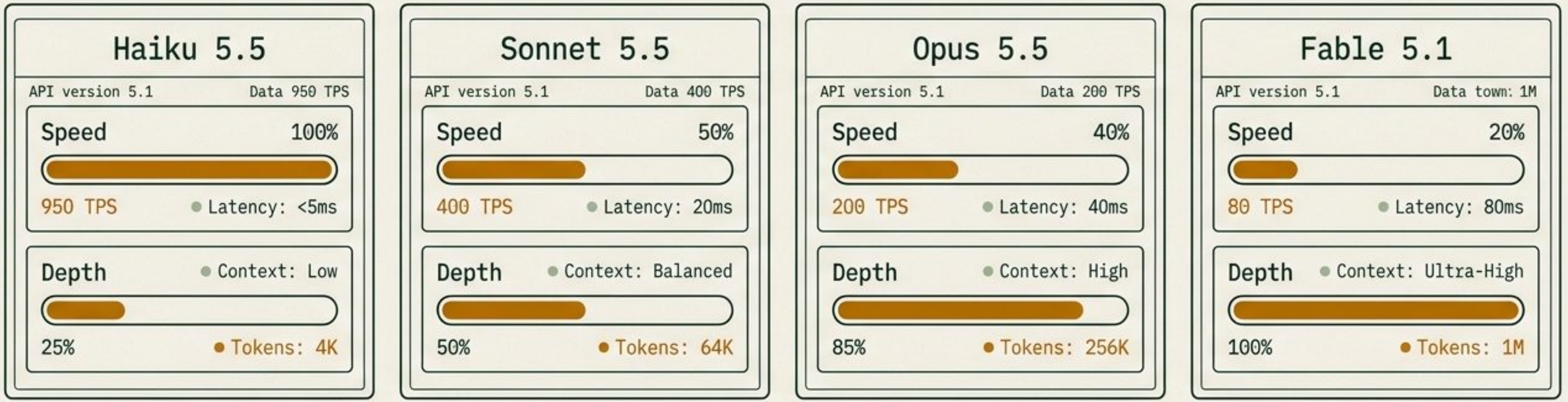
Four Brains, One Family

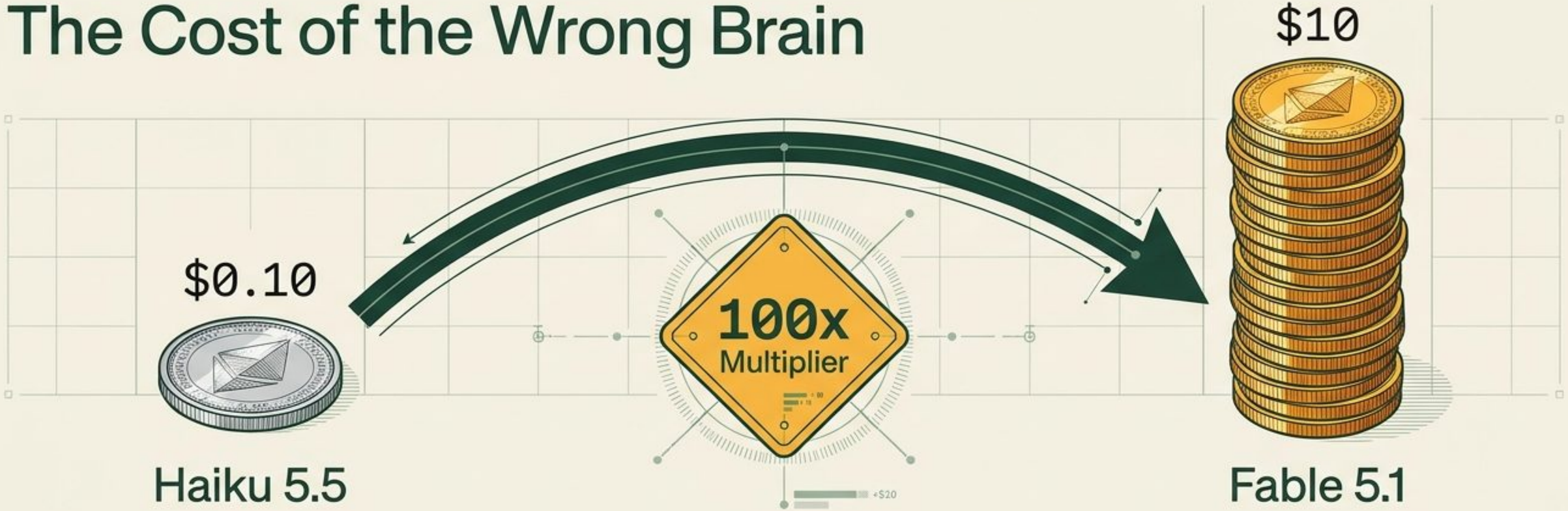
The definitive guide to orchestrating Claude 5.5 and 5.1 models.
Match the right mind to the right task.



© Copyright. All cognitive dashboard slide

Gemini Notebook

The Cost of the Wrong Brain



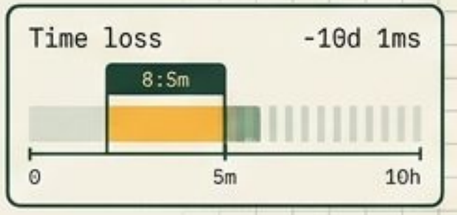
Wasting Money

Running routine text processing on the highest-tier model costs 100x more than necessary.



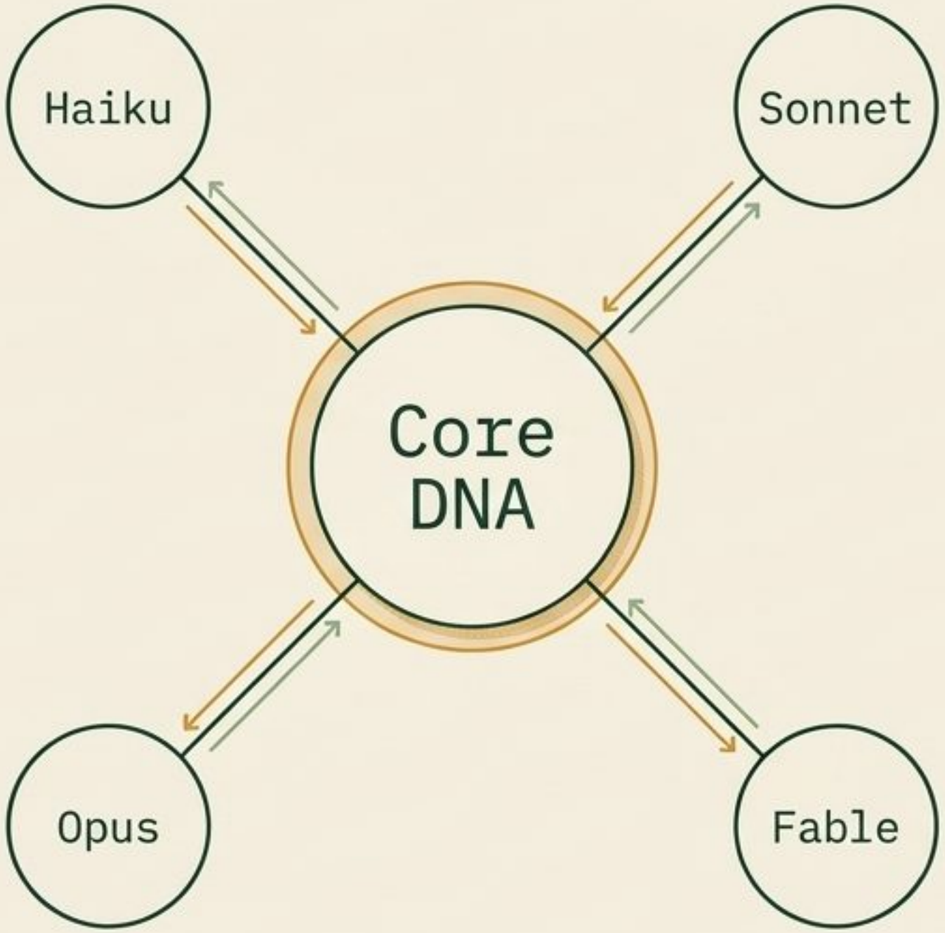
Wasting Time

Asking the fastest model to architect a complex application guarantees hours of manual fixes.

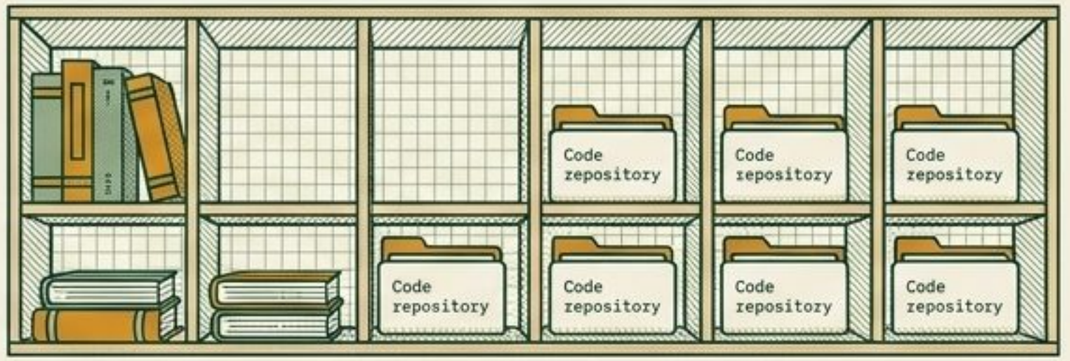


Gemini Notebook

The Baseline Physics



Context 1 Million Tokens



Output 128,000 Tokens

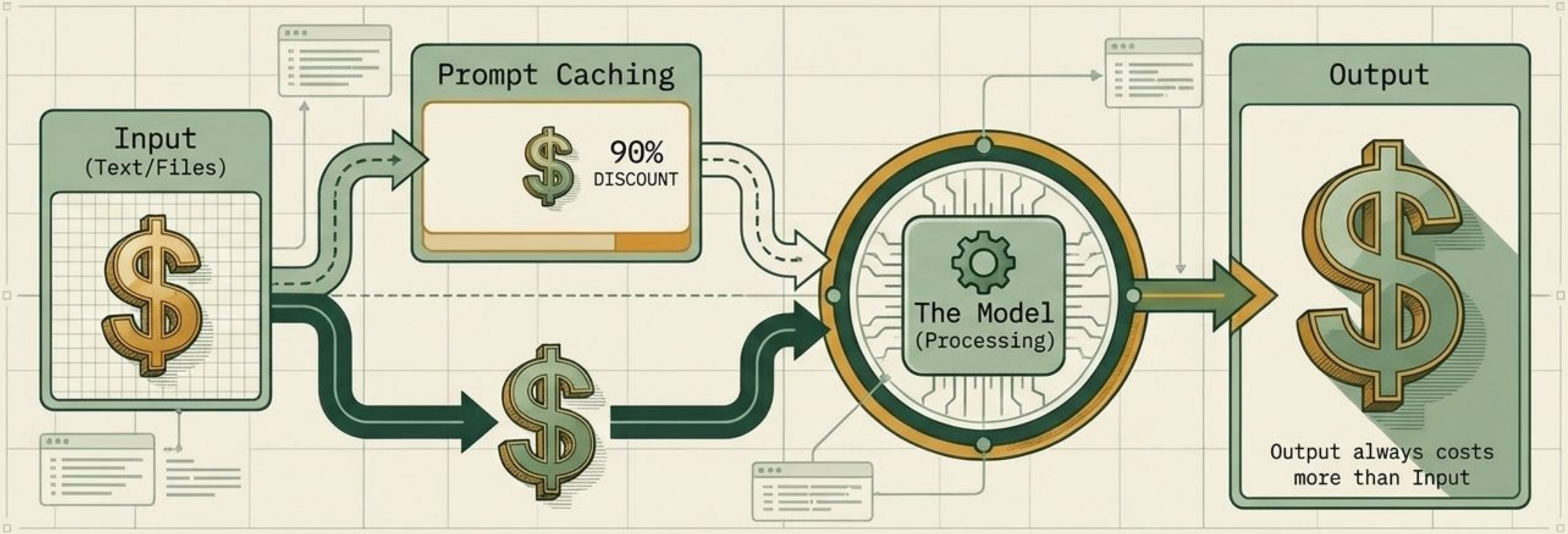


Knowledge June 2026 Cutoff



Gemini Notebook

How the Bill is Calculated

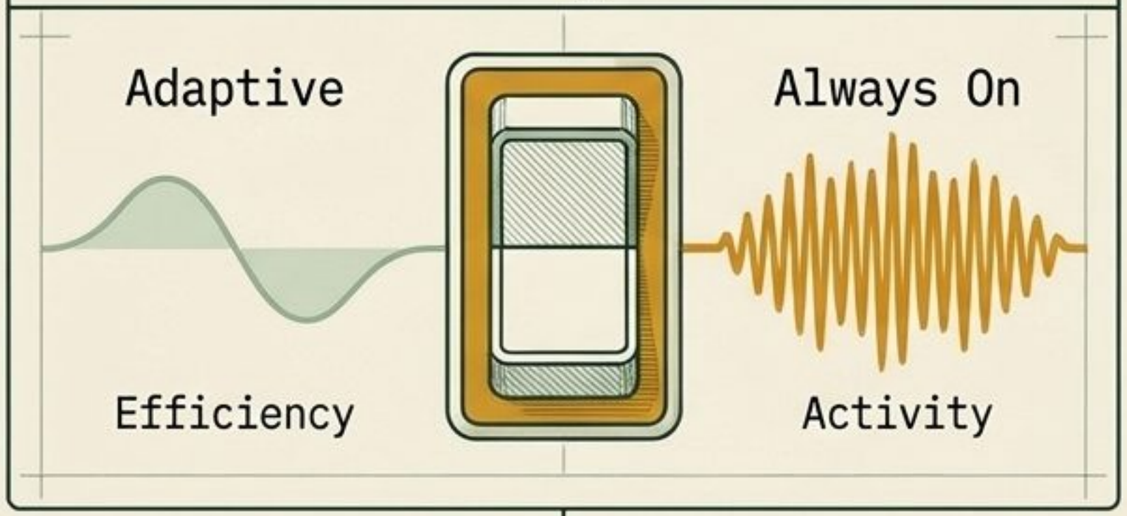


The Batch API Advantage: Send asynchronous jobs that don't need instant replies and receive a flat 50% discount on both input and output.

Gemini Notebook

The Hidden Levers: Thinking & Effort

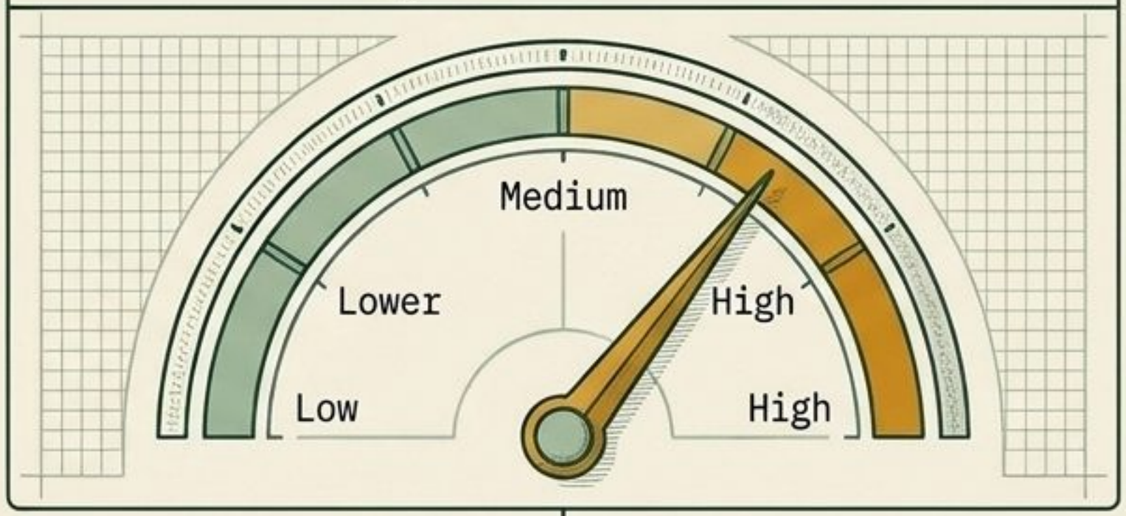
Thinking Mode



Fable & Opus

Locked into Always On thinking. Forced to to reason deeply before answering, maximizing accuracy at the expense of speed.

Cognitive Effort

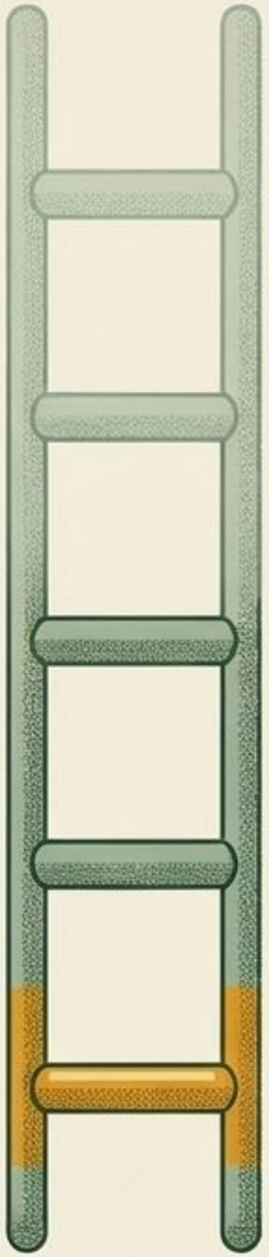


Sonnet & Haiku

Adaptive thinking. Bypasses deep reasoning for simple prompts to deliver near-instant replies.

Note: Haiku 5.5 is the first in its tier to feature an adjustable effort setting, granting unprecedented budget control.

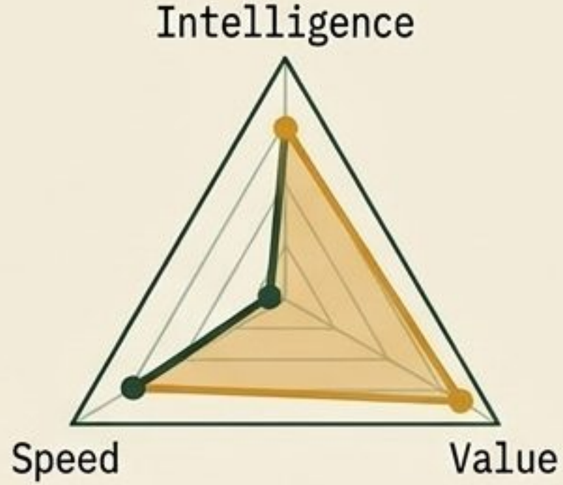
Gemini Notebook



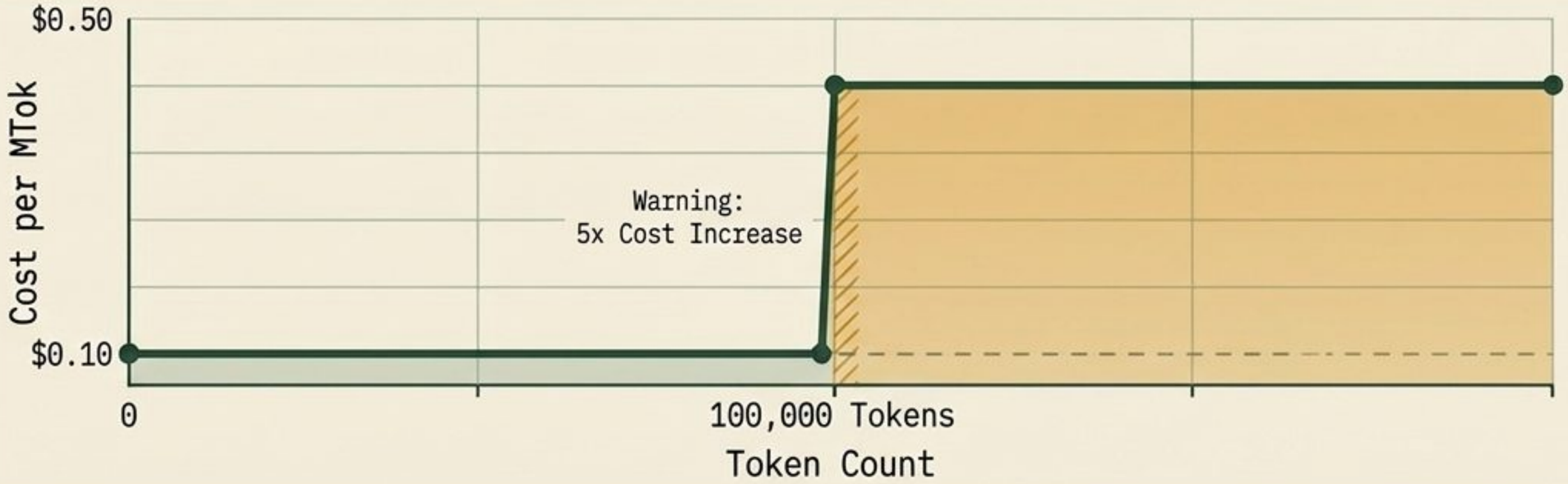
Haiku 5.5: The Speed King

$0.10 / MTok (Input)	$0.01 / MTok (Cache) / Medium	Adaptive Thinking Default Effort

Classification | Chatbots | High-Volume Sub-agents

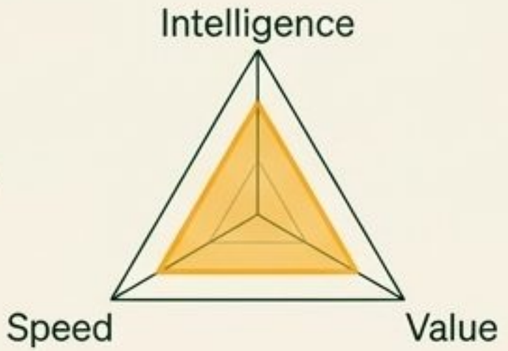


The 100K Trap



Gemini Notebook

Sonnet 5.5: The Balanced Workhorse



`$2.00 / MTok (Input) | $0.10 / MTok (Cache) | Adaptive Thinking / High Default Effort`

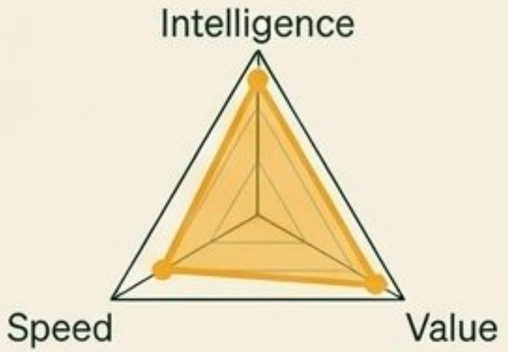
Daily Writing | Everyday Coding | PDF Analysis | Content Calendars

The Daily Driver

Included on the Free plan. Start every new task here. Move down to Haiku for volume; move up to Opus for complexity.

Gemini Notebook

Opus 5.5: The Everyday Flagship



`$4.00 / MTok (Input) | $0.20 / MTok (Cache) | Always On Thinking / Medium Default Effort`

Agentic Coding | App Architecture | Long Research Reports | Sub-agent Delegation

40%

40% cheaper to run than Opus 5.

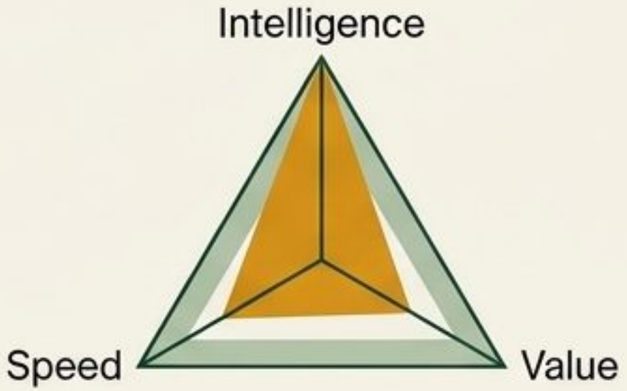
30%

30% faster output generation.

Matches Opus 5 agentic coding quality in half the time and tokens.

Gemini Notebook

Fable 5.1: Mythos-Class Power



`$10.00 / MTok (Input) | $0.25 / MTok (Cache) | Always On Thinking / High Default Effort`

Multi-day Software | Scientific Analysis | Complex Financial Models

The Twin Architecture

Fable 5.1

Generally available, embedded watermarking, 52.6% on Terminal Bench Science 0.1.

Mythos 5.1

Restricted vetted access only (cybersecurity/life sciences), identical underlying reasoning engine.

Gemini Notebook

The Executive Scorecard

Model	Intelligence (1-10)	Speed (1-10)	Value (1-10)	Access (1-10)	Target Audience
Fable 5.1	10	7	7.5	7	Researchers/ Enterprises
Opus 5.5	9.5	8.5	9.5	9	Developers/ Founders
Sonnet 5.5	9	9.5	9.5	10	Creators/ Professionals
Haiku 5.5	7.5	10	10	10	Support Teams/Volume

Gemini Notebook

System Architecture Specifications

Metric	Fable 5.1	Opus 5.5	Sonnet 5.5	Haiku 5.5
API ID	`claude-fable-5-1`	`claude-opus-5-5`	`claude-sonnet-5-5`	`claude-haiku-5-5`
Input/Output (per MTok)	$10/$50	$4/$20	$2/$10	$0.10/$0.50 (up to 100K)
Cache Read (per MTok)	$0.25	$0.20	$0.10	$0.01 (up to 100K)
App Plan	Pro/Max/Ultra	Pro+	All Plans	All Plans
Context / Max Out	1M / 128K	1M / 128K	1M / 128K	1M / 128K

Gemini Notebook

Real-World Economics

Job 1:
Typical Request
(20K in / 2K out)

Fable:	$0.30
Opus:	$0.12
Sonnet:	$0.06
Sonnet:	$0.06
Haiku:	$0.003

Job 2: Festive
Catalogue Bulk
(500K in / 300K out)

Fable:	**$20.00**
Opus:	$8.00
Sonnet:	$4.00
Sonnet:	$4.00
Haiku:	**$0.20**

Job 3:
Codebase Review
(400K in / 20K out)

Fable:	$5.00
Opus:	$2.00
Sonnet:	$1.00
Sonnet:	$1.00
Haiku:	$0.25

Gemini Notebook

The Secret: Hybrid Architecture

Top Funnel (Volume)

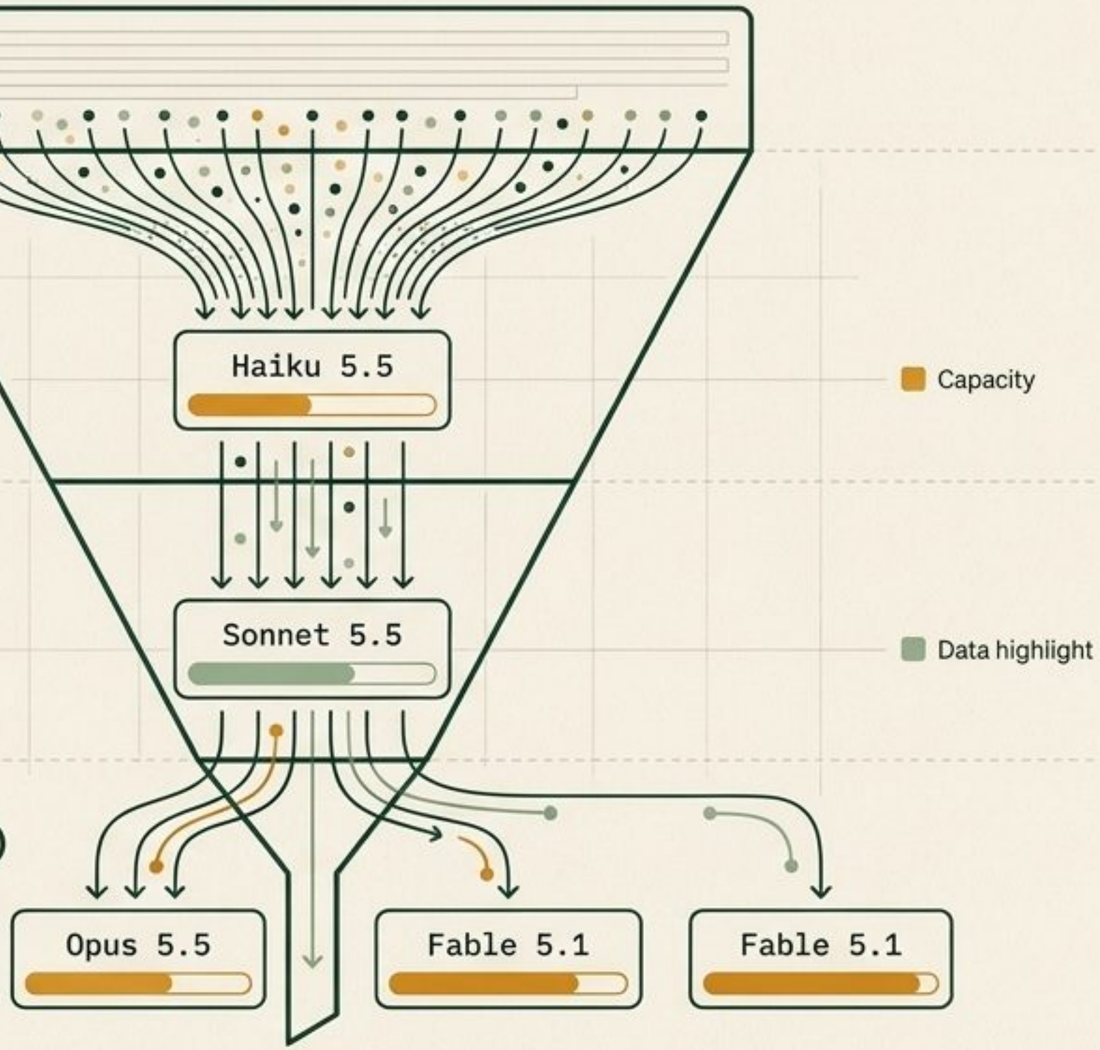
10,000 Raw Customer Reviews enter. **Haiku 5.5** sorts and tags them instantly for pennies.

Mid Funnel (Bulk Work)

Routine complaints are routed to **Sonnet 5.5** to draft standard, high-quality replies.

Bottom Funnel (Escalations)

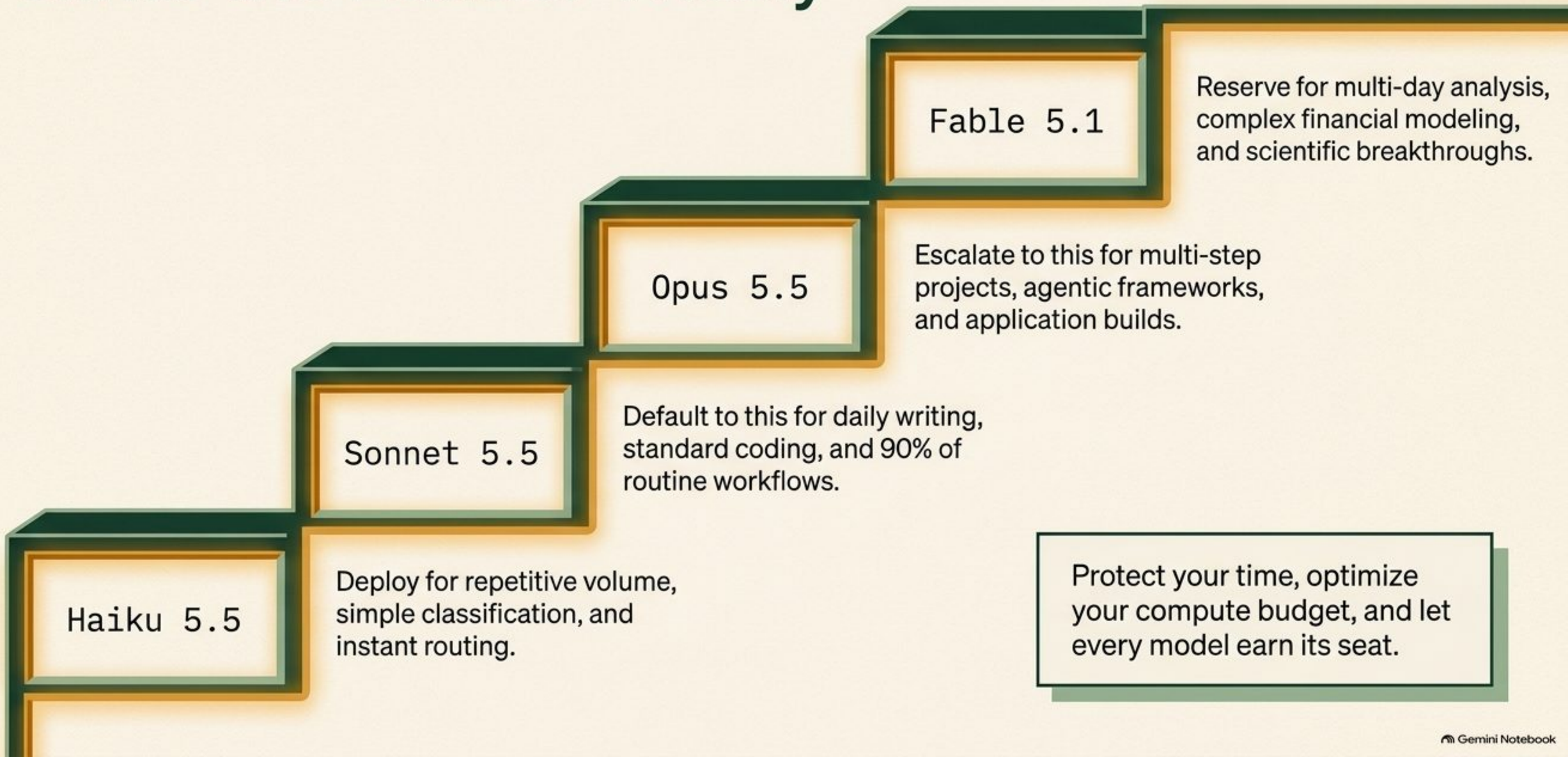
Complex legal or technical anomalies are routed to **Opus 5.5** or **Fable 5.1** for deep, resource-intensive reasoning.



Key Takeaway: Pay flagship prices only where flagship thinking pays off.

Gemini Notebook

Climb the Ladder Wisely



Protect your time, optimize your compute budget, and let every model earn its seat.

Gemini Notebook