

The frontier premium now lasts three weeks

On 22 September Anthropic released Claude Opus 5.5 at \$4 per million input tokens and \$20 per million output, and said it performs at the level of Claude Fable 5.1 on most work — a model the same company had launched 21 days earlier at \$10/\$50. The previous cycle, Fable 5 to Opus 5, took 45 days. When a lab's own next release caps the price of its flagship within weeks, the flagship stops being the thing that pays back frontier R&D and becomes the teacher for the tier that does.

AUTHOR



Alicante Tech Vibes *Research Labs*

~8 MIN READ
DATED

ALL FIGURES SOURCED &

\$4 / \$20 OPUS 5.5 PRICE PER MILLION INPUT / OUTPUT TOKENS · ANTHROPIC, 22 SEP 2026	21 days FROM FABLE 5.1 LAUNCH TO FABLE-LEVEL OPUS 5.5 · ANTHROPIC, SEP 2026	2.5× FABLE 5.1 PER-TOKEN PREMIUM OVER OPUS 5.5 · ANTHROPIC PRICE LISTS, SEP 2026
58 vs 53 INDEX SCORE AT MAX EFFORT, OPUS 5.5 VS FABLE 5.1 · ARTIFICIAL ANALYSIS, SEP 2026	~11% FABLE 5 SHARE OF ANTHROPIC BUSINESS SPEND · RAMP DATA VIA FT, AUG 2026	4 months GPT-5 LIFESPAN BEFORE REPLACEMENT; R&D NOT RECOUPED · EPOCH AI, JAN 2026

I. Anthropic undercut its own flagship in 21 days

On 22 September Anthropic released [Claude Opus 5.5](#) at \$4 per million input tokens and \$20 per million output, down from \$5/\$25 for Opus 5, with cache reads cut from \$0.50 to \$0.20. Anthropic says the model performs at Claude Fable 5.1 level on most work, costs about 40% less than Opus 5 on typical workloads and generates output more than 30% faster. It is generally available on the Claude Platform, AWS, Google Cloud and Microsoft Azure.

The comparison that matters is not with Opus 5. Fable 5.1, Anthropic's premium tier, [launched on 1 September](#) at \$10/\$50. Three weeks later the same company shipped a model it calls Fable-level at 40% of Fable's per-token price, with Fable-grade safeguards attached: [Unite.AI reports](#) three-stage cyber screening that falls back to Opus 4.8 and biology screening that falls back to Opus 5.

The premium tier's exclusive window is now measured in weeks. That changes what a frontier model is for: less a product that earns back its own R&D, more a teacher and a reference price for the tier that actually sells.

II. The lag between flagship and mainstream halved in one cycle

This is the second time this summer. Fable 5 [launched on 9 June](#) at \$10/\$50. Opus 5 followed on [24 July](#) at half the price, and Anthropic said it “approaches Fable-level performance at half the cost”. [The Register reported](#) Artificial Analysis putting Opus 5 one index point ahead of Fable 5 at launch. That cycle took 45 days. Fable 5.1 to Opus 5.5 took 21.

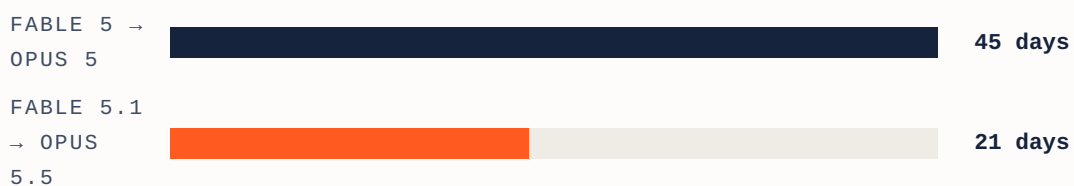
Fable 5's first window was shorter still in practice. US export controls imposed on 12 June restricted access for foreign nationals until Anthropic said [they were lifted on 30 June](#), leaving barely three weeks of unrestricted sale before its cheaper sibling arrived.

Buyers had already voted. Ramp data on roughly 70,000 US businesses, [reported by the Financial Times on 24 August](#), showed Fable 5 plateauing at about 11% of spending on Anthropic tools, with Opus 5 overtaking it. Ramp economist Ara Kharazian [attributed the shortfall](#) partly to OpenAI's GPT-5.6,

priced about 20% below Fable 5. Companies reserved the premium tier for the hardest jobs and routed the rest down.

Set that against how fast capability leaks out of a lab. Epoch AI estimates that the best open-weight models have trailed the closed frontier by about four months on average since January 2026. Anthropic's internal diffusion, from its flagship to its own cheaper tier, now runs roughly five to six times faster than external diffusion to open weights. And Opus 5.5 ships with an anti-distillation safeguard on its preserved reasoning. The lab is distilling its own frontier quickly while making it harder for anyone else to.

FIG. 1: DAYS FROM PREMIUM-TIER LAUNCH TO A FABLE-LEVEL OPUS RELEASE



BARS SCALED TO THE LONGER INTERVAL. RELEASE DATES: FABLE 5, 9 JUN 2026; OPUS 5, 24 JUL 2026; FABLE 5.1, 1 SEP 2026; OPUS 5.5, 22 SEP 2026. SOURCES: ANTHROPIC; VENTUREBEAT.

III. The honest technical read

Independent measurement supports the headline at maximum effort and blurs it at defaults. Artificial Analysis scores Opus 5.5 at 58 on its Intelligence Index against 53 for Fable 5.1, both at max effort, and measures Opus faster at 92 output tokens per second against 66. At each model's API default, Lindy's reading of the same data has both at 51, with Opus costing about \$1.34 a task against Fable's \$3.91.

The skeptic's case has three parts. Fable still leads in places: Lindy finds it ahead on GDPval-AA professional work at defaults (1617 against 1576 Elo) and better suited to long unattended agent runs. Token efficiency cuts the other way at max effort, where Opus burns roughly 119,000 output tokens per task against Fable's 78,000, so part of the per-token discount is spent on extra thinking. And Anthropic itself, as Lindy reports, says the gap in Opus's favour is “narrower than these scores suggest”, while Unite.AI quotes the company conceding that benchmark margins are now a less reliable guide at this level.

The 40% figure also needs unpacking. List prices fell 20% against Opus 5. The rest of Anthropic's claim rests on its own measurement of token use on typical workloads, which no independent source we found has replicated. Its

showcase example, an HAProxy translation finished in 9.5 hours instead of 12 at 51% lower cost, is a vendor anecdote, not a benchmark.

Safety review is the other caveat. Zvi Mowshowitz's [reading of the system card](#) calls Opus 5.5 in effect a smaller version of Fable 5.1, flags weaker refusals when harmful requests are split across conversations, and highlights METR's preliminary estimate of roughly 1.5x AI-driven acceleration in capabilities work. Unite.AI separately notes the system card lists regressions, including reader acceptance of unverifiable authorisation claims. If Fable-level capability now reaches the mass tier in three weeks, each distillate gets about three weeks of scrutiny too.

FIG. 2: COST PER TASK AT API DEFAULTS



BARS SCALED TO THE HIGHER COST. BOTH MODELS SCORE 51 ON THE INTELLIGENCE INDEX AT DEFAULTS (OPUS MEDIUM, FABLE HIGH). SOURCE: LINDY ANALYSIS OF ARTIFICIAL ANALYSIS DATA, 30 SEP 2026.

FIG. 3: OUTPUT TOKENS PER TASK AT MAX EFFORT — THE COUNTER-SIGNAL



BARS SCALED TO THE LARGER VALUE. AT MAX EFFORT OPUS 5.5 THINKS LONGER, ERODING PART OF ITS PER-TOKEN DISCOUNT. SOURCE: LINDY ANALYSIS OF ARTIFICIAL ANALYSIS DATA, 30 SEP 2026.

When a lab's own next release sets the ceiling on its flagship's price within three weeks, pricing power at the top is bounded by the release calendar, not by competitors.

FABLE 5.1 LAUNCHED AT \$10/\$50 ON 1 SEPTEMBER 2026. OPUS 5.5, DESCRIBED BY ANTHROPIC AS FABLE-LEVEL ON MOST WORK, LAUNCHED AT \$4/\$20 ON 22 SEPTEMBER.

IV. R&D payback moves down a tier

Prices. Fable-level capability fell 60% in per-token price in 21 days, and roughly 66% in cost per task at defaults on Lindy's numbers. Pricing power at the top of the stack is now bounded by the lab's own release cadence. A

buyer who can wait three to six weeks gets the flagship's capability at mass-tier prices.

Capital. Epoch AI's [model-lifecycle analysis](#) found GPT-5 earned about \$6 billion over a four-month life with roughly \$2 billion of gross profit, against about \$5 billion of R&D in the preceding months: it did not pay for itself, at gross margins Epoch puts near 30%. A premium window of three weeks makes that arithmetic far harsher. Illustrative estimate, not a sourced figure: take Anthropic's reported [\\$65 billion run rate](#) (Bloomberg via Fortune, 18 August) and apply Ramp's ~11% Fable share to get roughly \$7 billion a year of premium-tier revenue; 21 days of that is about \$0.4 billion, or around \$0.12 billion of gross profit at Epoch's margin. The assumptions are generous — Ramp's US business sample stands in for all revenue, and it ignores that Fable keeps some customers after Opus ships — but the order of magnitude holds. A three-week premium pays for a rounding error of a frontier programme. Payback has to come from volume at the Opus tier, which is exactly where Anthropic is cutting prices.

Labour. At about \$1.34 a task for Fable-level work, model cost stops being what limits automation of complex software and knowledge work; review, verification and liability are. Anthropic cites a tester completing a 680,000-line code migration in under a day. That is a vendor anecdote, but it shows where substitution lands first: senior engineering work that was gated by price is now gated by trust.

Market structure. Fast internal distillation is a scale economy. A rival without its own frontier model waits for external diffusion, about four months on Epoch's measure; a lab that distils its own flagship in three weeks takes the volume tier first and protects the source with anti-distillation measures. The durable asset is not the frontier model itself but the speed of turning it into something cheap.

V. Second-order: the public premium tier hollows out

If the public premium tier cannot hold a price, its role shifts to things a cheaper tier cannot offer. Anthropic already sells one: [Mythos 5.1](#), the same model as Fable 5.1 with looser safeguards, restricted to vetted cybersecurity and life-sciences organisations. Opus 5.5 routes most cyber tasks to Opus 4.8

unless the customer joins the Cyber Verification Program. What stays scarce is permission, not intelligence.

Buyers will also learn to wait. If Fable-level capability reaches Opus prices within weeks, procurement teams defer premium commitments, which shortens the premium's revenue window further. Fable 5's 11% plateau suggests that behaviour started before Opus 5.5 arrived.

The third effect falls on safety review. Each fast distillate is a new deployment with its own system card, and the review window shrinks with the release interval. The regressions Anthropic itself disclosed are the cost of that pace made visible.

VI. How it plays out: three scenarios

Probabilities are our analytical judgment, not model forecasts, and they sum to 100%.

1 · Baseline: the flagship becomes a teacher ~55%

6-18 MONTHS ·
STEADY

Anthropic keeps shipping a premium tier followed by a Fable-level Opus within one to two months. Fable holds a niche in long-running agents and gated Mythos access; Opus-tier prices drift down each cycle. Watch for Fable 5.1 holding \$10/\$50 through year-end while its Ramp spend share stays near or below 11%, the next Opus following its Fable by under 60 days, and rivals adopting the same flagship-then-distillate pattern.

2 · Acceleration: the lag goes to zero ~25%

3-9 MONTHS · FAST

The next premium release ships alongside its own distillate, or the public premium tier is retired in favour of a gated Mythos-style tier plus a mass Opus tier. Frontier R&D is funded entirely from volume and enterprise contracts. Watch for a Fable price cut before January, a same-week flagship and Opus launch, and OpenAI or Google matching their flagship at a cheaper tier within a month.

3 · Stall: the gap reopens ~20%

6-12 MONTHS · ABRUPT

A larger capability jump or tighter bio and cyber gating stops Fable-level capability reaching the mass tier quickly, and premium pricing power returns. Watch for more than 90 days between the next Fable release and a matching Opus, Fable's share of Ramp spend rising above 20%, and regulators or AI safety institutes requiring pre-release review for each distilled model.

WHAT WOULD FALSIFY THIS THESIS

The claim is that the premium tier's exclusive window is short and shrinking, and that frontier R&D payback has moved to the volume tier. It fails if the next Fable-class release goes more than 90 days without a Fable-level Opus, or if production evidence shows Opus 5.5 well behind Fable 5.1 on real agentic work — visible as Fable's share of business spend rising rather than falling after 22 September. It also fails if Anthropic's IPO disclosures show the premium tier carrying a disproportionate share of gross profit despite its small share of spend.

WHAT TO WATCH, IN ORDER

1. **By 31 October 2026** — Ramp's next spend read-out: whether Opus 5.5 pushes Fable 5.1 below Fable 5's ~11% share.
2. **By 31 December 2026** — Fable 5.1 list price: a cut below \$10/\$50 means Anthropic concedes pricing power at the top tier.
3. **Q4 2026** — Cyber Verification Program expansion, which Anthropic says is coming soon: how much Fable-grade cyber capability Opus 5.5 unlocks for verified users.
4. **Anthropic IPO prospectus, expected late 2026** — the first audited view of gross margin and R&D intensity, and whether payback really sits in the volume tier.
5. **Next Fable-class release, likely H1 2027** — days until a Fable-level Opus ships. Under 21 confirms the trend; over 90 supports the stall case.

SOURCES

1. Anthropic, "Claude Opus 5.5," 22 September 2026 · anthropic.com
2. MacRumors, "Anthropic Launches Claude Opus 5.5 With Fable-Level Performance at a Lower Price," 22 September 2026 · macrumors.com
3. Unite.AI, "Anthropic Releases Claude Opus 5.5 With Lower Pricing and New Safeguards," 22 September 2026 · unite.ai
4. VentureBeat, "Anthropic's Claude Fable 5.1 and Mythos 5.1 arrive with a 75% cost reduction for Fable cache reads," 1 September 2026 · venturebeat.com
5. Anthropic, "Introducing Claude Opus 5," 24 July 2026 · anthropic.com
6. The Register, "Anthropic debuts Opus 5 at half the price of its Fable sibling," 25 July 2026 · theregister.com
7. Anthropic, "Redeploying Claude Fable 5," 30 June 2026 · anthropic.com
8. Artificial Analysis, "Claude Opus 5.5 vs Claude Fable 5.1," accessed 4 October 2026 · artificialanalysis.ai
9. Lindy, "Opus 5.5 vs Fable 5.1: Same Score at a Third of the Cost," 30 September 2026 · lindy.ai
10. Zvi Mowshowitz, "Claude Opus 5.5: The System Card," 23 September 2026 · thezvi.wordpress.com
11. Techmeme / Financial Times (George Hammond), Ramp data on Fable 5 spending share, 24 August 2026 · techmeme.com
12. Parameter, "Business Users Turn Away From Anthropic's Premium AI Model in Favor of Budget Options," 25 August 2026 · parameter.io
13. Epoch AI, "Can AI companies become profitable?" 28 January 2026 (updated 6 March 2026) · epoch.ai
14. Epoch AI, "Open models lag state-of-the-art closed models by 4 months," 29 May 2026 · epoch.ai
15. Fortune, "Anthropic's annual revenue run rate reportedly hits \$65 billion," 18 August 2026 · fortune.com

METHOD NOTE: FIGURES ARE ATTRIBUTED AND DATED INLINE; WHERE SOURCES DISAGREE, THE DISAGREEMENT IS REPORTED RATHER THAN AVERAGED. BENCHMARK AND COST-PER-TASK FIGURES ARE ARTIFICIAL ANALYSIS MEASUREMENTS, AS PUBLISHED BY ARTIFICIAL ANALYSIS OR ANALYSED BY LINDY, NOT VENDOR CLAIMS; ANTHROPIC'S 40% SAVING AND CUSTOMER ANECDOTES ARE LABELLED AS VENDOR CLAIMS. ARTIFICIAL ANALYSIS HAS RE-BASED ITS INDEX SINCE JULY, SO SCORES FROM DIFFERENT MONTHS ARE NOT COMPARED. THE PREMIUM-WINDOW REVENUE FIGURE IS AN ILLUSTRATIVE ESTIMATE WITH ITS ASSUMPTIONS STATED. SCENARIO PROBABILITIES ARE ANALYTICAL JUDGMENT.