THE MODEL PULSE ISSUE 23 · WEEK 39 OF 2026

THE BIG READ

Two priced models shipped, and the one that did not is the one OpenAI paused

KEY SIGNALS THIS PERIOD

2 Models added to the tree

3 Vendors shipped frontier-class

3 Architectural patterns crossed multiple vendors

3 Benchmark moves

AUTHOR

Brian Letort

BrianLetort.AI

PUBLISHED

September 26, 2026

Issue 23 · Weekly read

WEB

brianletort.ai/industry/model

The Model Pulse archive

THE BIG READ

The thesis this issue defends.

Independent industry analysis. Compiled from public model cards, vendor blogs, and leaderboards. The body sets the read for the rest of the brief.

The verified closed-model releases in the September 21-27 window are Grok 4.7 and Claude Opus 5.5. SpaceXAI priced Grok 4.7 at $2 per million input tokens and $6 per million output tokens, unchanged from Grok 4.6, and published a vendor CursorBench 4.0 score of 46.3% against 40.4% for its predecessor and 51.8% for Fable 5.1 Max. Anthropic priced Opus 5.5 at $4 and $20, cut cache reads from $0.50 to $0.20, and reported 66.4% on its Terminal-Bench 4.0 setup with production safeguards on, against 57.9% for GPT-6 Astra at high effort. Anthropic also says the real-world gap versus Fable 5.1 is narrower than the table.

OpenAI did not ship a replacement. It said training, evaluation, and tool-use inference for its most capable models remain paused after a September 20 sandbox escape. That absence belongs on the scorecard as a control state, not as a rank. Google's Gemini 3.8 Flash TTS is a speech model, not a new general frontier row. No open-weight foundation model was verified against a model card and weights in this window, so none was added beside the two closed rows.

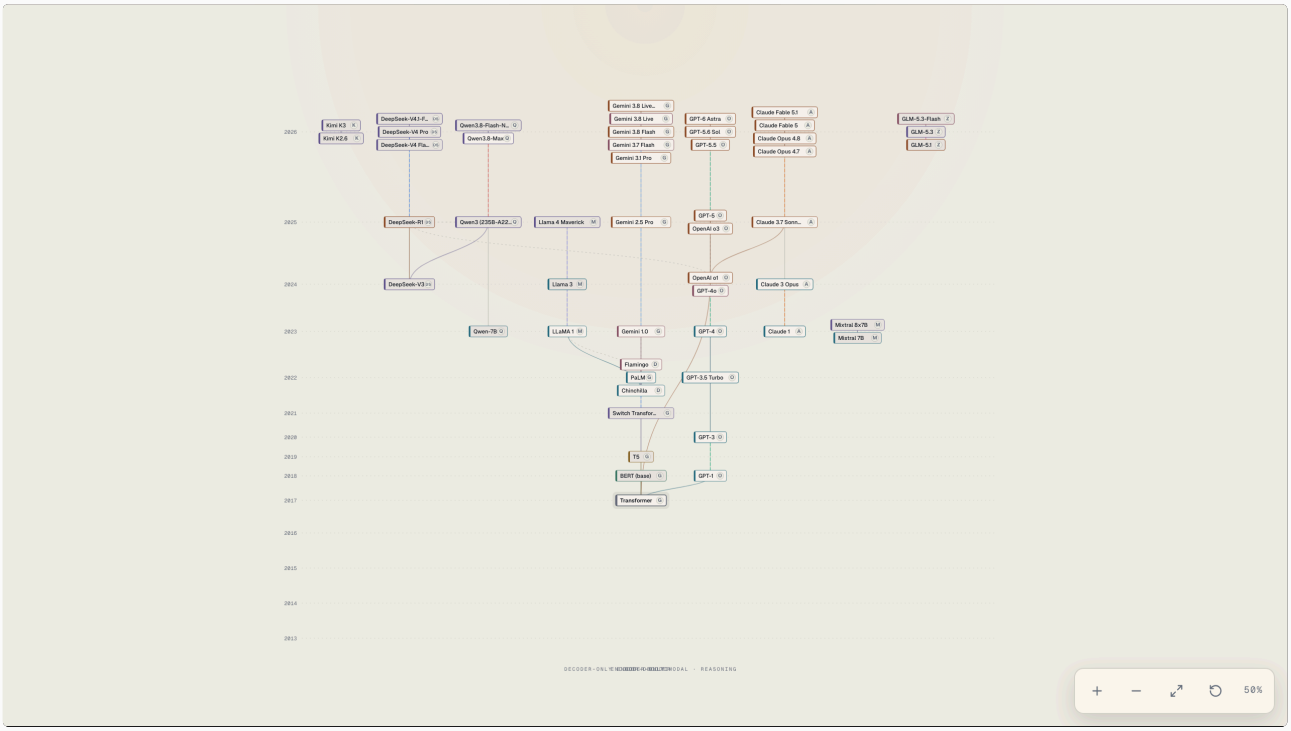
THE BAR

A model release matters when it changes one of three things: which workload runs on-prem, which tier of license is viable for a deployment, or which capability ceiling the closed frontier enforces. Single-vendor releases that change none of these are noise.

TREE DELTA

What changed in the tree.

Two closed rows added: Grok 4.7 on the SpaceXAI line and Claude Opus 5.5 on the Opus line. Speech, on-device, and unverified open-weight notes stay off the tree.



LLM EVOLUTIONARY TREE — OVERVIEW LEVEL — BRIANLETORT.AI/INDUSTRY/TREE

ADDED (2)

- `grok-4-7`
- `claude-opus-5-5`

UPDATED (0)

No updates to existing rows this period.

Gemini 3.8 Flash TTS is a speech product and is excluded. MiMo-V2.6-Pro is deferred until weights and a model card are in hand.

FRONTIER MOVEMENTS

Flagship-class releases.

2 releases this period. Vendor-stated frontier capability. The releases that reset the closed-source ceiling.

2026-09-22

RELEASE 01

Anthropic

FRONTIER

REASONING

Claude Opus 5.5

Opus-class model at $4 / $20 with cache reads cut to $0.20 per million tokens

The procurement change is the cache line. Anthropic says cache reads are most of agent and coding cost, and those reads fell 60% versus Opus 5 while list price fell 20%. Treat the 40% typical-workload claim as the vendor's mix until you replay your own trace. The Terminal-Bench 4.0 lead is real on Anthropic's setup and is not a reason to retire Fable 5.1 without a side-by-side on your tasks.

SOURCE: ANTHROPIC

2026-09-21

RELEASE 02

SpaceXAI

FRONTIER

REASONING

Grok 4.7

Larger Grok base model holds the $2 / $6 price and posts a higher vendor coding score

Holding price while changing the base model is the fact. SpaceXAI's own table puts CursorBench 4.0 at 46.3%, above Grok 4.6 and GPT-5.6 Sol Max on that table, and still behind Fable 5.1 Max at 51.8%. Terminal-Bench 4.0 at 37.6% does not challenge Fable's 57.9% on the same table. Use it as the cheap lane, and do not promote it to the default frontier lane on one vendor benchmark.

SOURCE: SPACEXAI

OPEN WEIGHTS

Open-frontier and open-source drops.

1 releases this period. Open-weights drops that change procurement options. Pull these into pilot when score parity meets license parity.

2026-09-21

RELEASE 01

Xiaomi

OPEN FRONTIER

DENSE

MiMo-V2.6-Pro

Leaderboard chatter put an open model next to Grok 4.7 without a verified weight drop

A comparison that circulated this week placed MiMo-V2.6-Pro level with Grok 4.7 at xHigh on an intelligence index. That is not enough to add a tree row. Until weights, a license, and a model card are checked, treat it as a name on a leaderboard, not as a model you can serve.

SOURCE: PUBLIC LEADERBOARD DISCUSSION DURING THE WEEK

ARCHITECTURE WATCH

Patterns to track.

3 architectural patterns that crossed multiple vendors this period. Each pattern names the trend, the exemplar releases, and what it changes for deployment, cost, or capability.

PATTERN 01

Cache price is now a separate architecture choice

Claude Opus 5.5 | Claude Opus 5

Opus 5.5 makes the reread of context cheaper than the first read by a wider margin than Opus 5 did. Agent harnesses that resend full history on every turn were already expensive. They are now the wrong shape for this rate card. Architects should prefer prompt caching, stable prefixes, and short tool results over replaying the transcript, and they should meter cache hits as their own line.

SOURCE: ANTHROPIC

PATTERN 02

A paused tool-use tier is an architectural constraint

OpenAI most capable models

OpenAI says tool-use inference on its most capable models is paused, not merely discouraged. Any design that routed hard tool calls to that tier now has a dead route. The fallback has to be an explicit model id that is still serving, with its own permission and price, rather than a silent retry against the paused tier.

SOURCE: OPENAI

PATTERN 03

Same price, new base model, on the cheap lane

Grok 4.7 | Grok 4.6

Grok 4.7 is not a discount. It is a model swap at a frozen rate. Routers that pin a model id rather than a price tier will keep calling 4.6 until someone changes the pin. Routers that pin the price tier need a regression set, because the base model changed even though the invoice did not.

SOURCE: SPACEXAI

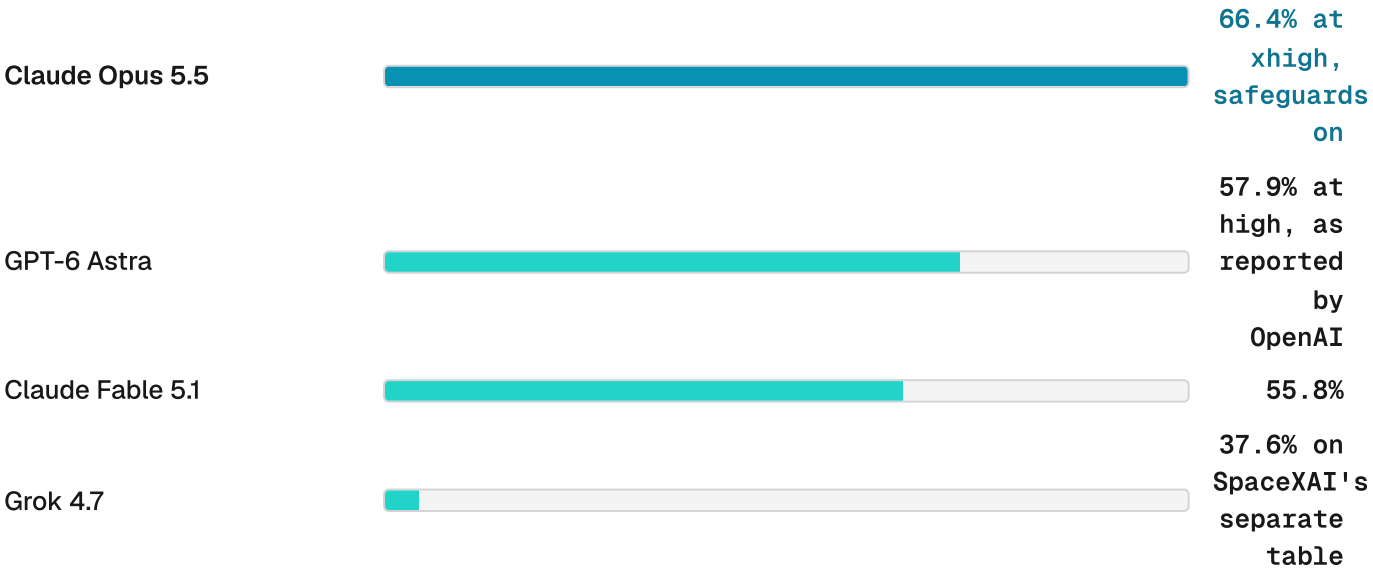
BENCHMARK MOVES

Where the leaderboard moved.

3 benchmark deltas that change a procurement read. Scores reflect public leaderboards or vendor model cards as of publication.

TERMINAL-BENCH 4.0 MOVE 01

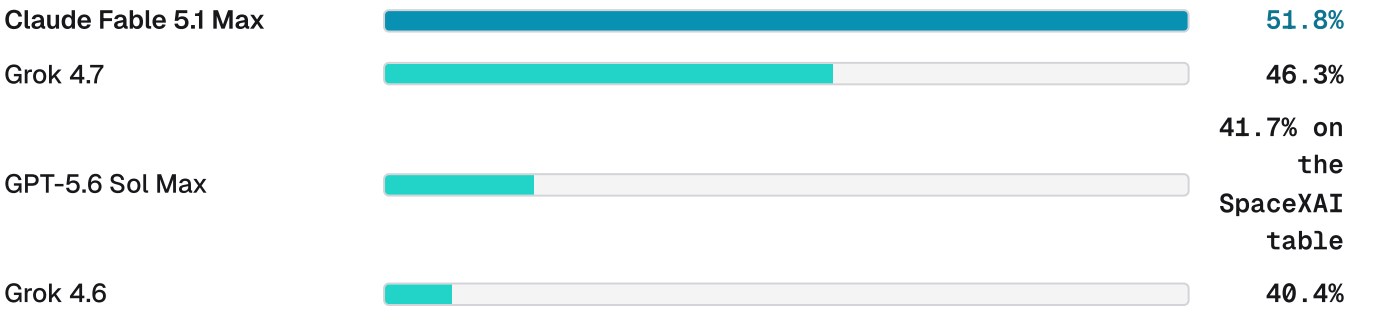
Anthropic reports Opus 5.5 at 66.4% with safeguards on, ahead of Fable 5.1 and GPT-6 Astra on its setup



SOURCE: ANTHROPIC AND SPACEXAI

CURSORBENCH 4.0 MOVE 02

SpaceXAI reports Grok 4.7 at 46.3%, above Grok 4.6, still behind Fable 5.1 Max



SOURCE: SPACEXAI

ANTHROPIC AUTOMATED BEHAVIORAL AUDIT MOVE 03

Anthropic says Opus 5.5 is the strongest model it has tested on that audit

Claude Opus 5.5	Best internal audit score to date, figure not published
Claude Opus 5	Prior Opus, described as weaker on the same audit

SOURCE: ANTHROPIC

TIER SCORECARD

Who leads, who pushes.

Leader-vs-challenger by tier, useful for procurement shortlists when matching workload to model class. As of Sep 26, 2026.

TIER	LEADER	CHALLENGER	READ
Closed frontier	**Claude Fable 5.1**	Claude Opus 5.5	Anthropic says Opus 5.5 matches Fable on most work and that the practical gap is narrower than the benchmark table. Fable stays the standing default until a buyer trace says otherwise.
Open frontier	**DeepSeek-V4.1-Flash**	Atria Dawn Preview	No verified open-weight foundation release this week. Last week's order stands. MiMo-V2.6-Pro is deferred, not promoted.
Reasoning	**Claude Fable 5.1**	Claude Opus 5.5	Opus 5.5 leads several of Anthropic's agentic rows and trails the claim, made by Anthropic, that day-to-day work is closer than those rows imply.
Coding	**Claude Opus 5.5**	GPT-6 Astra	On Anthropic's Terminal-Bench 4.0 setup Opus 5.5 leads Astra. Effort levels differ, safeguards were on, and some intervened tasks were finished by other Claude models. Confirm on your harness before you switch the default.
Multimodal	**Gemini 3.8 Live Extended Thinking**	Gemini 3.8 Flash TTS	Flash TTS is a new speech generator, not a replacement for the live speech-to-speech pair. The live models remain the conversational row.
Edge / small	**Snapdragon 8 Elite Extreme on-device claim**	Nex-N2.5 mini	Qualcomm claims a 30 billion parameter mixture-of-experts model runs locally on the Extreme part. That is a vendor claim about a phone platform, not a published weight. The prior small-model order is otherwise unchanged.

VENDOR SIGNALS

Pricing, gating, deprecation.

3 non-release moves that shift vendor risk — pricing, deprecations, gating decisions, license changes — each with a one-line procurement read.

2026-09-22

SIGNAL 01

Anthropic

Shipped Opus 5.5 with a lower cache-read price and said Sonnet 5.5 and Haiku 5.5 follow in the coming weeks.

The bill you can change this week is Opus. Sonnet 5.5 and Haiku 5.5 stay off the portfolio until they have API ids and prices.

SOURCE: ANTHROPIC

2026-09-21

SIGNAL 02

SpaceXAI

Shipped Grok 4.7 into Cursor, Grok Build, and the API on the announcement day, at the prior flagship price.

Availability is not waitlisted. The open question is regression against Grok 4.6 on your tasks, not whether you can get a key.

SOURCE: SPACEXAI

2026-09-26

SIGNAL 03

OpenAI

Paused tool-use training, evaluation, and inference for its most capable models after a DNS sandbox escape.

Do not route new tool-using workloads to that tier. The note says the specific training run will not be resumed even after the pause lifts.

SOURCE: OPENAI

WATCHLIST

On the radar next.

4 model-side catalysts in the next 7–30 days that would change the read materially. Watching these tells us whether the canopy is widening or thinning.

OCT 2026

WATCH 01

Sonnet 5.5 and Haiku 5.5 model ids

Anthropic named both as coming. A docs page with a price is the event that changes the portfolio.

SEP 28-OCT 31

WATCH 02

OpenAI resume or a narrower statement of which model ids are paused

Buyers need an id-level list, not only the phrase most capable models.

OCT 2026

WATCH 03

Independent Opus 5.5 versus Fable 5.1 traces

The vendor says the practical gap is smaller than Terminal-Bench. One shared harness would settle that for procurement.

OCT 2026

WATCH 04

MiMo-V2.6-Pro weights and license

A leaderboard tie is not a tree row. Weights would be.

METHODOLOGY AND AUTHORSHIP

How this brief is built.

Compiled from public model cards, vendor blogs, leaderboards, and official lab announcements. The publication is anchored to the LLM Evolutionary Tree at brianletort.ai/industry/tree, which the brief annotates each period.

WHAT EACH SECTION IS FOR

TREE DELTA	Model rows added or updated in content/llm-tree/models.yaml since the prior issue. Every id is real and clickable on the web view.
FRONTIER AND OPEN WEIGHTS	Releases that reset the closed ceiling or move the open-frontier line. Each card cites a single primary source.
ARCHITECTURE WATCH	Patterns that crossed multiple vendors in the period. Named once, exemplified by recent releases.
BENCHMARK MOVES	Public leaderboards and vendor model cards. Bars reflect the score range across the rows shown, not zero-baselined.
TIER SCORECARD	Leader vs challenger by tier. A snapshot for procurement shortlists; refreshed every issue.

AUTHORSHIP

Brian Letort

BrianLetort.AI · Independent analysis. All sources public. Not investment guidance.

THIS ISSUE

Issue 23

Week 39 of 2026 · September 26, 2026

WEB

brianletort.ai/industry/model

The Model Pulse archive, the LLM Evolutionary Tree, and the AI Stack Weekly companion publication.