

THE BIG READ

# April rewrote the floor: open weights crossed the closed frontier, and the safety-gated frontier became a procurement criterion.

KEY SIGNALS THIS PERIOD

11

Models added to the tree

6

Vendors shipped frontier-class

4

Architectural patterns crossed multiple vendors

3

Benchmark moves

AUTHOR

**Brian Letort**  
BrianLetort.AI

PUBLISHED

**April 24, 2026**  
Issue 01 · Monthly recap

WEB

**[brianletort.ai/industry/model](https://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.

April 2026 was the most productive month in frontier model history. Eleven model rows landed in the LLM Evolutionary Tree across eight vendors, including the first open-weights MoE to score frontier-class on SWE-Bench Pro coding (Kimi K2.6 at 58.6 and GLM-5.1 at 58.4, both above GPT-5.4 at 57.7 and Claude Opus 4.6 at 57.3). DeepSeek V4 shipped in two configurations under MIT license with 1M-token context and 1.6T total parameters, moving on-prem coding from 'can we?' to 'which workload first?' Anthropic withheld its Mythos flagship on cyber-capability grounds after UK AISI confirmed autonomous offensive capability — an inflection that turns capability gating from a research-org concern into a procurement diligence requirement. The closed frontier responded: GPT-5.5 (Apr 23) and Claude Opus 4.7 (Apr 16) reset the closed-source ceiling, with adaptive-thinking and ultra-long-context as the new defaults. Read together, April was the month the canopy widened on three axes at once: open-vs-closed parity on coding, reasoning-as-default at every tier, and capability gating as a market signal.

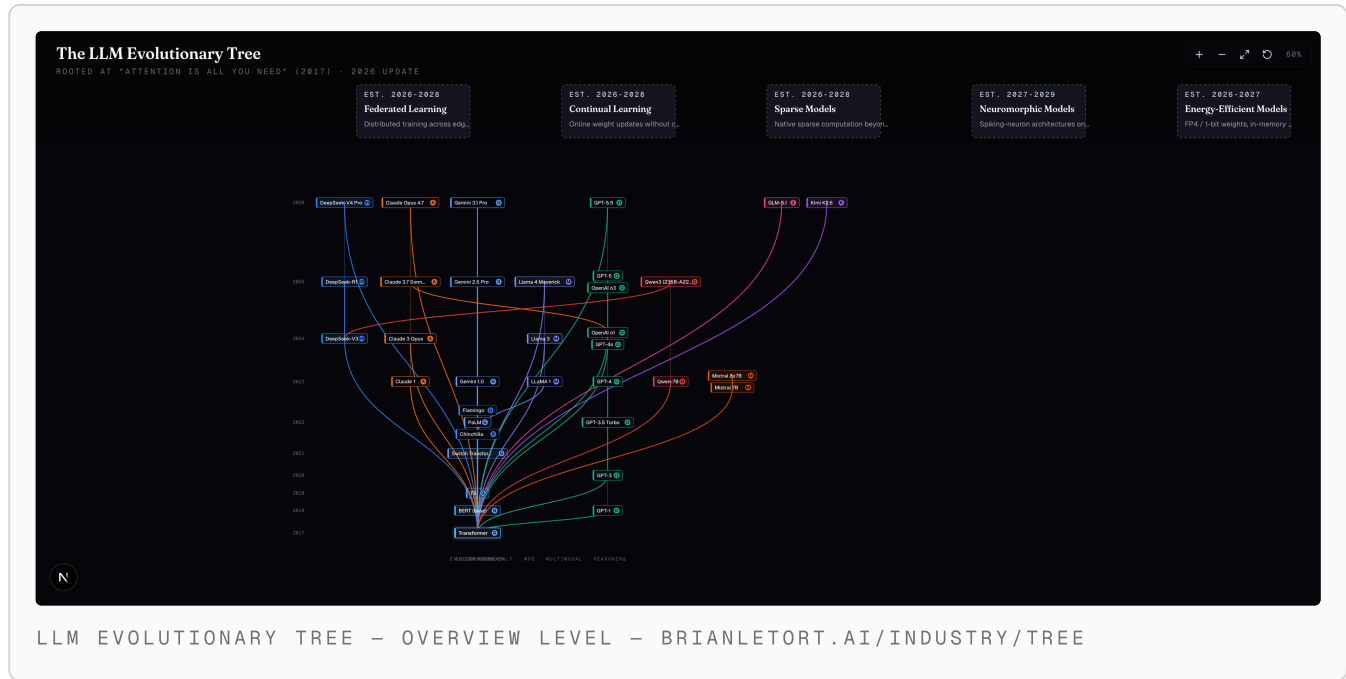
**THE BAR**

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.

11 model rows added to the tree in April, spanning frontier closed releases, open-frontier MoEs, gated safety-class previews, and reasoning-tier successors. The decoder-only zone widened most; the multimodal and reasoning branches absorbed every major release.



ADDED ( 11 )

- claude-opus-4-6
- qwen-3-6-plus
- claude-mythos
- glm-5-1
- claude-opus-4-7
- qwen-3-6-35b-a3b
- kimi-k2-6
- gpt-5-5
- deepseek-v4-pro
- deepseek-v4-flash
- deepseek-r2

UPDATED ( 4 )

- claude-opus-4-5
- kimi-k2-5
- gpt-5-4
- qwen-3-5

*The Anthropic line gained two flagship releases plus one gated preview in a single month. DeepSeek shipped V4 in two scales (Pro and Flash) plus a same-month R2 reasoning successor — a cadence no other vendor matched in April.*

## FRONTIER MOVEMENTS

# Flagship-class releases.

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

**2026-04-16**

RELEASE 01

Anthropic

FRONTIER

AGENTIC

## Claude Opus 4.7

**Adaptive-thinking flagship with ultra-long context and full agentic stack (computer-use + MCP + tool use).**

Reset the closed-frontier reasoning ceiling. The agentic surface area (computer-use, MCP, tool-use) makes Opus 4.7 the default reference for closed-frontier procurement when adaptive thinking and long-horizon agent runs are required. Pair with Sonnet 4.6 for the cost-tiered deployment.

SOURCE: ANTHROPIC.COM RELEASE NOTES, MODEL CARD

**2026-04-23**

RELEASE 02

OpenAI

FRONTIER

REASONING

## GPT-5.5

**Frontier reasoning flagship with multimodal and ultra-long-context as defaults.**

OpenAI's response to the V4 / Opus 4.7 surge. Sets the closed-frontier reasoning bar that the open-frontier challengers (V4 Pro, K2.6, GLM-5.1) measure against on SWE-Bench Pro and GPQA. The 5.x family now spans Instant, Thinking, Pro, mini, and nano tiers — useful for matching workload class to model class.

SOURCE: OPENAI.COM BLOG, GPT-5.5 SYSTEM CARD

**2026-04-07**

RELEASE 03

Anthropic

SPECIALIST

AGENTIC

## Claude Mythos Preview

**Frontier-class capability withheld from public access on cyber-capability grounds.**

The first major model gated by its own developer for safety reasons after independent (UK AISI) red-teaming confirmed autonomous offensive capability. Vendor-risk frameworks now need a capability-gate criterion alongside availability SLAs — diligence on 'what could the next release do that the current one cannot?' is no longer optional.

SOURCE: ANTHROPIC.COM, RED.ANTHROPIC.COM, AISI.GOV.UK

**2026-04-02**

RELEASE 04

Anthropic

FRONTIER

REASONING

## Claude Opus 4.6

**Predecessor to 4.7; held the SWE-Bench Pro closed-frontier line at 57.3 mid-month.**

Useful as the April 02 baseline against which open-weights eventually moved past mid-month. Procurement teams running 4.5 should plan upgrade to 4.7 directly; 4.6 will likely move to maintenance status before mid-May.

SOURCE: ANTHROPIC.COM RELEASE NOTES

## OPEN WEIGHTS

# Open-frontier and open-source drops.

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

**2026-04-24**

RELEASE 01

DeepSeek AI

OPEN FRONTIER

MOE

## DeepSeek-V4 Pro

**1.6T MoE, 1M context, MIT license, frontier-class on coding.**

The headline release of April. First open-weights model with frontier-class SWE-Bench Pro performance under a permissive license. Procurement teams blocked on closed-source data residency or licensing constraints now have an open frontier alternative for coding workloads. The on-prem floor moved up; the merchant-vs-self-host fork is now real.

SOURCE: [HUGGINGFACE.CO/DEEPSEEK](https://huggingface.co/deepseek), [DEEPSEEK.COM](https://deepseek.com)**2026-04-24**

RELEASE 02

DeepSeek AI

OPEN FRONTIER

MOE

## DeepSeek-V4 Flash

**Cost-tiered V4 sibling for high-throughput inference on smaller fabric.**

Open-weights answer to GPT-4o-mini and Claude Sonnet at the inference-cost tier. Useful for batch coding agents and high-volume serving where Pro is overkill. Same MIT license, same 1M context, lower active-parameter cost per token.

SOURCE: [HUGGINGFACE.CO/DEEPSEEK](https://huggingface.co/deepseek) MODEL CARD**2026-04-20**

RELEASE 03

Moonshot AI

OPEN FRONTIER

MOE

## Kimi K2.6

**58.6 on SWE-Bench Pro — the first open-weights model to top the closed leaders on coding.**

The benchmark moment. K2.6's coding score sits above GPT-5.4 (57.7) and Claude Opus 4.6 (57.3) under an open-weights license. For coding-heavy procurement reads, K2.6 is the new default open-frontier baseline; closed-source premium needs new justification beyond raw score.

SOURCE: [MOONSHOT.CN](https://moonshot.cn), [ARTIFICIAL ANALYSIS](https://artificialanalysis.ai) SWE-BENCH PRO**2026-04-08**

RELEASE 04

Z.AI (Zhipu)

OPEN FRONTIER

AGENTIC

## GLM-5.1

**58.4 on SWE-Bench Pro with multi-agent-native architecture and ultra-long context.**

Second open-weights model to clear the closed-frontier coding bar this month. The multi-agent-native design point makes GLM-5.1 the open-source reference for agent-of-agents deployments — relevant when coordination overhead matters more than single-shot completion.

SOURCE: [Z.AI](https://z.ai), [GLM-5.1 MODEL CARD](https://glm-5.1.modelcard.ai)

**2026-04-16**

RELEASE 05

Alibaba

OPEN FRONTIER

MOE

## Qwen3.6-35B-A3B

**Open-weights MoE successor to Qwen3.5; agentic + multimodal at deployable scale.**

The most deployable of April's open releases on commodity inference fabric. 35B total / 3B active makes Qwen3.6-A3B the open multimodal pick when V4 Pro is too large and K2.6 is single-modality. Pairs naturally with Qwen3.6-Plus on the closed-source side for hybrid deployments.

SOURCE: QWENLM.GITHUB.IO, MODEL CARD

**2026-04**

RELEASE 06

DeepSeek AI

REASONING

REASONING

## DeepSeek-R2

**Open-weights reasoning successor to R1, frontier-class long-context.**

Same vendor shipping V4 (general) and R2 (reasoning) inside one calendar month is the cadence story. R2 closes the open-vs-closed reasoning gap that R1 opened in 2025 — open-weights reasoning is now a separate procurement category, not a niche.

SOURCE: DEEPSEEK.COM, HUGGINGFACE.CO/DEEPSEEK

## ARCHITECTURE WATCH

# Patterns to track.

4 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 Reasoning becomes the default mode, not a separate model.

Claude Opus 4.7 (adaptive thinking)

GPT-5.5

DeepSeek-R2

GLM-5.1

April's flagships ship reasoning behavior as a default rather than a separate o-series-style fork. Adaptive thinking, where the model decides how much to deliberate, is now the closed-frontier expectation; explicit reasoning toggles look dated by month-end. Procurement consequence: the reasoning-vs-non-reasoning split is collapsing — assume reasoning is on, budget tokens accordingly.

SOURCE: VENDOR MODEL CARDS (ANTHROPIC, OPENAI, DEEPSEEK, Z.AI)

## PATTERN 02 Million-token context as baseline.

DeepSeek-V4 Pro (1M)

Kimi K2.6 (1M+)

Claude Opus 4.7 (ultra-long)

GPT-5.5 (ultra-long)

Every flagship released in April clears the 1M-token threshold. Long-context is no longer a differentiator at the frontier; it is the floor. Workloads that previously required RAG architectures may now be expressible as direct in-context loads — re-evaluate the retrieval layer when re-baselining for V4 / K2.6 / Opus 4.7.

SOURCE: MODEL CARDS; HUGGINGFACE TECHNICAL REPORTS

## PATTERN 03 Agentic surface area as a first-class capability.

Claude Opus 4.7 (computer-use, MCP, tool-use)

GLM-5.1 (multi-agent-native)

Qwen3.6-Plus (agentic)

Kimi K2.6 (agentic-ready)

Agentic isn't a benchmark category anymore; it is a model property declared on the card. Computer-use, MCP support, and multi-agent coordination ship with the flagship rather than as a downstream wrapper. Architectural read: the boundary between 'model' and 'agent runtime' is dissolving — procurement should evaluate the full agent surface, not just the LM weights.

SOURCE: VENDOR MODEL CARDS; MODEL CONTEXT PROTOCOL ANNOUNCEMENTS

## PATTERN 04

**Capability gating as a market signal.**

Claude Mythos Preview (withheld)

Anthropic withholding Mythos after UK AISI red-team findings is the first time a major lab gated a frontier-class release on its own initiative for cyber-capability reasons. The signal: capability evaluation now happens before public release, and 'gated' is a status that procurement teams need to track in vendor risk frameworks alongside 'available' and 'deprecated.'

SOURCE: ANTHROPIC.COM, RED.ANTHROPIC.COM, AISI.GOV.UK

---

BENCHMARK MOVES

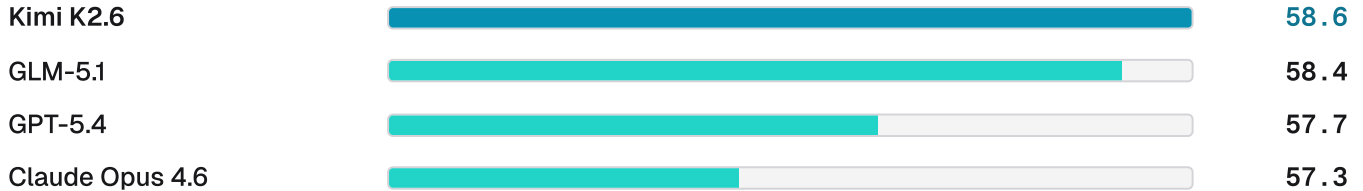
# Where the leaderboard moved.

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

## SWE-BENCH PRO (CODING)

MOVE 01

**Open weights overtook closed for the first time. Two open-weights models clear both closed leaders.**



SOURCE: ARTIFICIAL ANALYSIS SWE-BENCH PRO LEADERBOARD, VENDOR MODEL CARDS

## LONG-CONTEXT RETRIEVAL (1M+ TOKENS, VENDOR-REPORTED)

MOVE 02

**Million-token context lands at the frontier with sub-2% retrieval-error rates across multiple vendors.**

DeepSeek-V4 Pro	1M ctx, 1.4% err
Kimi K2.6	1M ctx, 1.7% err
Claude Opus 4.7	ultra-long, vendor-reported
GPT-5.5	ultra-long, vendor-reported

SOURCE: VENDOR TECHNICAL REPORTS; HUGGINGFACE EVALUATIONS

## AGENTIC TOOL-USE (VENDOR + THIRD-PARTY)

MOVE 03

**Closed flagships still lead agentic tool-use; open frontier is one tier behind but closing.**

Claude Opus 4.7	Closed leader
GPT-5.5	Closed challenger
GLM-5.1	Open leader (multi-agent-native)
DeepSeek-V4 Pro	Open challenger

SOURCE: VENDOR AGENT BENCHMARKS; COMMUNITY EVALUATIONS

## TIER SCORECARD

# Who leads, who pushes.

Leader-vs-challenger by tier, useful for procurement shortlists when matching workload to model class. As of Apr 24, 2026.

TIER	LEADER	CHALLENGER	READ
<b>Closed frontier</b>	<b>Claude Opus 4.7</b>	GPT-5.5	Adaptive thinking is the closed-source differentiator this month.
<b>Open frontier</b>	<b>DeepSeek-V4 Pro</b>	Kimi K2.6	MIT license + 1M context + frontier coding score is the new open baseline.
<b>Reasoning</b>	<b>DeepSeek-R2</b>	GPT-5.5 (thinking mode)	Open-weights reasoning is now its own procurement category.
<b>Coding</b>	<b>Kimi K2.6</b>	GLM-5.1	Open-weights tops both closed leaders on SWE-Bench Pro this month.
<b>Multimodal</b>	<b>Gemini 3.1 Pro</b>	Qwen3.6-Plus	Closed leader from Q1 still holds; Qwen3.6 narrows the gap on the open side.
<b>Edge / small</b>	<b>Phi-4-mini</b>	Gemma 4	Q1 leader carries through April; no major edge-tier release this month.

## VENDOR SIGNALS

# Pricing, gating, deprecation.

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

**2026-04-21**

SIGNAL 01

Anthropic

## Mythos withheld from public access pending capability-gate review.

Sets a precedent for vendor-initiated gating on cyber-capability grounds. Procurement frameworks now need a 'gated' status alongside 'available' / 'deprecated.' Adds a new diligence question: what is the next release the vendor evaluated and chose not to ship?

SOURCE: ANTHROPIC.COM, RED.ANTHROPIC.COM, AISI.GOV.UK

**2026-04-24**

SIGNAL 02

DeepSeek AI

## Two-config V4 launch (Pro + Flash) plus same-month R2 release under MIT.

Cadence and licensing both shifted. A monthly cycle with permissive licensing across coding, reasoning, and inference-tier variants is the new open-frontier benchmark for vendor velocity. Closed-source vendors with quarterly cadence have a velocity gap, not just a capability gap.

SOURCE: DEEPSEEK.COM RELEASE NOTES; HUGGINGFACE.CO MODEL CARDS

**2026-04**

SIGNAL 03

OpenAI

## GPT-4 family deprecation timeline tightened; 5.x consolidated to five tiers (Instant, Thinking, Pro, mini, nano).

Pricing migration for GPT-4-era workloads accelerates through Q2. Customers should plan migration to 5.x within 60 days; the tier consolidation simplifies model selection but compresses the price-vs-capability slope on the low end.

SOURCE: OPENAI.COM PLATFORM CHANGELOG

**2026-04**

SIGNAL 04

Multiple (open-weights serving)

## Inference price cuts of 25-40% across major open-weights serving providers following V4 / K2.6 launches.

Open-weights inference economics improved materially this month. Re-baseline cost per token for batch agentic workloads against the new floor; the merchant-vs-self-host break-even moved.

SOURCE: PROVIDER PRICING PAGES; COMMUNITY COMPARISON THREADS

## WATCHLIST

## On the radar next.

5 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.

**MAY 1 - 7**

WATCH 01

**Closed-frontier response to V4 / K2.6 coding scores.**

Anthropic and OpenAI both have credible reasons to push a coding-tuned point release to recover the SWE-Bench Pro lead. Watch for Opus 4.7-Codex or GPT-5.5-Codex within two weeks.

**MAY 1 - 14**

WATCH 02

**Hyperscaler Q1 prints (MSFT / GOOG / META / AMZN).**

AI capex commentary will set the inference-pricing trajectory for the rest of Q2. Tracked in The AI Stack Weekly's capital-flow lens; relevant here as the upstream signal for serving costs.

**MAY 5 - 20**

WATCH 03

**Mythos status update from Anthropic + AISI.**

If gating becomes permanent or extended, 'gated' becomes a durable status class. If gating lifts under conditions, a capability-gate playbook emerges — either way, the procurement read changes.

**MAY 10 - 25**

WATCH 04

**Open-source reasoning successor cluster.**

DeepSeek-R2 set the bar; Qwen, Z.AI, and Moonshot all have R2-class candidates plausibly within 30 days. Watch for the second open-weights reasoning model to clear o-series parity.

**MAY 15 - 31**

WATCH 05

**First gated-capability standardization proposals.**

Mythos creates pressure for a cross-vendor gating taxonomy (capability levels, evaluation protocols, release criteria). Watch for AISI / METR / OpenAI / Anthropic joint statements.

## 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](https://brianletort.ai/industry/tree), which the brief annotates each period.

**WHAT EACH SECTION IS FOR**

<b>TREE DELTA</b>	Model rows added or updated in <code>content/llm-tree/models.yaml</code> since the prior issue. Every id is real and clickable on the web view.
<b>FRONTIER AND OPEN WEIGHTS</b>	Releases that reset the closed ceiling or move the open-frontier line. Each card cites a single primary source.
<b>ARCHITECTURE WATCH</b>	Patterns that crossed multiple vendors in the period. Named once, exemplified by recent releases.
<b>BENCHMARK MOVES</b>	Public leaderboards and vendor model cards. Bars reflect the score range across the rows shown, not zero-baselined.
<b>TIER SCORECARD</b>	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 01**

April 2026 Recap · April 24, 2026

**WEB****[brianletort.ai/industry/models](https://brianletort.ai/industry/models)**

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