

Is Qwen Going Closed? Open Weights vs Frontier (2026)

June 15, 2026


[Download this guide as PDF](#)

Quick Answer: Qwen split into two tracks, not one closed tier. The frontier (3.7-Max, 3.7-Plus, Qwen-VLA) is proprietary and API-only. The open mid-tier (3.6-27B and 35B-A3B) is still on Hugging Face under Apache 2.0, current as of April. What's missing: the 3.7 open 27B and 35B variants Alibaba announced. Four weeks past the 3.7-Max launch and no repo has appeared. This guide breaks down the split, explains why your 3.6 setup is still current, and names the one signal worth watching.

If you run Qwen locally, you've probably noticed a confusing question on r/LocalLLaMA over the last month: is Qwen going closed?

The short answer is no. The longer answer is more useful: Qwen split. Alibaba pushed the 3.7 generation into a closed, proprietary frontier tier – 3.7-Max, 3.7-Plus, the new robotics-focused VLA model – while keeping a current open mid-tier that runs roughly one generation behind on Hugging Face under Apache 2.0. Your 3.6-27B or 3.6-35B-A3B setup is still the current open Qwen.

What's unclear, and worth your attention if you run open weights, is whether the 3.7 open 27B and 35B variants Alibaba announced will actually ship, or whether the open tier is starting to lag the closed one in a more permanent way. Below is the honest read: what's confirmed, what's still genuinely uncertain, and what to actually do about it.

 **Live tracker:** [Qwen 3.7 Preview Scored 57 AAI: 27B/35B Open Weights Next](#) – that piece tracks the AAI score and the open-weight drop. This article is the strategic explainer of the split itself.

The Two Tracks

Closed Frontier: API-only, Proprietary

The 3.7 generation, so far, has shipped entirely as closed-weight, paid API.

Model	Released	What It Is	Pricing	Open Weights?
Qwen 3.7-Max	May 19, 2026 (public announce May 20)	Reasoning agent, 1M context, AAI Intelligence Index 56.6 (#5 overall, #1 Chinese model)	\$2.50 / \$7.50 per 1M tokens	No
Qwen 3.7-Plus	June 1, 2026	Multimodal agent (text, image, video in; text out), 1M context, ~6x cheaper than Max	\$0.40 / \$1.16 per 1M via Alibaba; \$0.32 / \$1.28 via OpenRouter	No
Qwen-VLA	May 29, 2026	Vision-language-action model for embodied AI and robotics (paper at arxiv 2605.30280)	Research preview	Proprietary

Three closed releases in under a month. That's a faster cadence than Qwen's open-tier releases have ever run at, and it represents a real strategic move by Alibaba: a paid frontier offering that competes head-on with Claude Opus, GPT-5.x, and Gemini. Qwen 3.7-Max specifically ranks #1 among Chinese models on Artificial Analysis's Intelligence Index, sitting at #5 overall against the current US-lab frontier.

Open Mid-Tier: Live on Hugging Face Now, Current as of April

While the 3.7 closed tier was launching, the 3.6 open tier sat right where Alibaba left it in April – current, runnable, and Apache 2.0.

Model	Released	Type	Repo	License
Qwen 3.6-27B	April 22, 2026	Dense, multimodal coder. 77.2 on SWE-bench Verified – the best open dense coder right now.	Qwen/Qwen3.6-27B	Apache 2.0
Qwen 3.6-35B-A3B	April 16, 2026	MoE, 35B total / 3B active per token. Sparse activation makes single-3090 inference viable.	Qwen/Qwen3.6-35B-A3B	Apache 2.0

These are roughly two months old. They're on Hugging Face, they run locally, and they're still the current open frontier from Alibaba. For most local AI workloads, 3.6-27B is the open backbone right now.

If you're running an older Qwen, the [Qwen 3.6 local guide](#) is your upgrade path. The broader family context is in the [Qwen Models Guide](#).

The Open Question: Where Are the 3.7 Open Weights?

Here's the part Alibaba hasn't resolved.

When the 3.6 family launched in April, the 27B dense and 35B-A3B MoE open-weight releases shipped on Hugging Face within about a week of the announcement. That's the recent baseline: Qwen announces, weights drop, the community runs benches inside 24 hours.

The 3.7 release has not followed that pattern. The 3.7-Max API went live May 19. The 3.7-Plus GA hit June 1. As of June 15, no Qwen3.7-27B or Qwen3.7-35B-A3B repository exists on Hugging Face under the official Qwen organization. Direct probes of the plausible repo names return HTTP 401 – Hugging Face's response for nonexistent repos. The 3.7 open weights have been announced as forthcoming. They haven't shipped.

Case for “delayed, not abandoned”:

- Qwen's open-weight cadence through 3.5 and 3.6 was extremely consistent. Alibaba ships open weights.
- The 3.7 open variants were officially announced. That's a real commitment from a team that has historically kept this kind of promise.
- Engineering a 27B dense plus a 35B-A3B MoE for open release after a Max-tier API launch is non-trivial work. A six-week lag isn't unreasonable for that scope.
- The 3.6 family itself shipped over a staggered one-week window (35B-A3B on April 16, 27B on April 22). The 3.7 equivalent could land on a similar staggered schedule that just hasn't started yet.

Case for “the gap is widening”:

- The June release window the 3.6 pattern predicted has now passed without a repo or a firmer date.
- Alibaba is building out the closed tier aggressively. Three closed releases in under a month, each with its own price tier and revenue logic, is a serious commercial push.
- The 3.7-Plus is a multimodal agent – a non-trivial commercial offering with real revenue potential. There's a real incentive against releasing open weights that would substitute for the API.
- Four weeks of silence after a major announcement is the kind of pattern that sometimes looks, in retrospect, like the start of a permanent gap rather than a delay.

I'm not going to pick. Anyone telling you they know which read is right is guessing. Both cases are live, both have real support, and the evidence is genuinely ambiguous. The signal that will

resolve the question is concrete: a Qwen3.7-* repo appearing on the official HF org, or continued silence past the next Alibaba Cloud milestone.

What This Means For You

If you're running open Qwen locally today, here's the practical move.

Stay on 3.6. The 3.6-27B and 35B-A3B weights are current, on Hugging Face, Apache 2.0, and aren't going anywhere. The 27B dense model is the best open coder at the moment by a fair margin, and the MoE variant lets you fit serious capability into a single 3090's VRAM. If you don't have a working 3.6 setup yet, the [3.6 local guide](#) is the place to start. There's no reason to wait on 3.7 open weights to settle your local rig, and given the genuine uncertainty above, "waiting" might be open-ended.

If you need the frontier today, it's API-only. Qwen 3.7-Max is API-only at \$2.50 / \$7.50 per million tokens. If your workload needs Max-tier reasoning, you can call it through Alibaba Cloud Model Studio, OpenRouter, Together AI, or Qubrid AI. The 3.7-Plus multimodal agent is the cheaper alternative if you don't need Max-tier depth — \$0.40 / \$1.16 per million tokens at Alibaba's rate, or \$0.32 / \$1.28 on OpenRouter. That's a real cost spread between the two closed tiers.

If your strategy depended on Qwen being fully open across all tiers, recalibrate. The split is real, and the closed frontier looks structural rather than temporary. Whether the open tier eventually catches up to 3.7, or stays one generation behind on a more permanent basis, is the open question. For now, plan around running open Qwen 3.6 locally for everything it covers, and budget for the API where you actually need frontier capability you can't run locally.

What to Watch

The single concrete signal worth tracking is the official Qwen organization on Hugging Face. If a Qwen3.7-27B or Qwen3.7-35B-A3B repository appears, that resolves the question in favor of the "delayed, not widening" read, and the open tier is back on the expected cadence. If we get to mid-July without one, the "gap is widening" interpretation starts to look more accurate.

The [Qwen 3.7 Preview / Open Weights Watch](#) article is the live tracker for this. InsiderLLM will publish firsthand RTX 3090 plus RTX 3060 benches within 24 hours of any open-weight release.

A secondary signal is what comes out of Alibaba next. If Qwen pushes another closed-tier variant before any 3.7 open weights drop — a 3.7-Nano, a 3.7-Coder, anything new behind a

paywall – that’s stronger evidence the closed side is the priority now. If the next thing out of Alibaba is the 3.7 open 27B, the inverse is true.

For now: 3.6 is current, the API is the frontier, and the open question is honestly open.

Get notified when we publish new guides.

[Subscribe – free, no spam](#)

Source: <https://insiderllm.com/guides/qwen-open-weights-vs-closed-frontier-2026/>

Free guides for running AI locally