

The Convergence Illusion

Open-weight models are closing the benchmark gap — enterprise adoption is moving the other way

By Diego Molina — Mediavertical

Open-weight models have closed most of the performance gap with closed frontier systems, but enterprise adoption of open models is falling, not rising — a signal that switching costs, not capability, now decide the buying decision.

The benchmark gap has genuinely narrowed

By mid-2026, 42% of the top 50 tracked AI models were open-weight, and the performance gap between the top closed and top open model on human-preference rankings had narrowed to roughly 28 points, down from about 56 points a year earlier. Open models now close an estimated 70–90% of the remaining capability gap on frontier benchmarks.

The economics have moved even faster than the benchmarks. Open models deliver that performance at roughly five to ten times lower cost per token than closed alternatives — a gap wide enough that cost alone should, in theory, be pulling workloads toward open weights.

Enterprise adoption has moved in the opposite direction

It hasn't. Open-source model adoption across enterprise workloads actually declined to 13% in 2026, down from 19% in early 2025, as enterprises consolidated around closed providers instead. Roughly 80% of token volume still flows to closed models, which cost on average about six times more per token than open alternatives.

That is the counterintuitive part of the story: the capability and cost arguments for open models have strengthened at exactly the moment enterprises pulled back from using them.

What is actually driving the closed-model premium

Enterprises appear to be paying for the system, not the model in isolation. Once workloads reach production scale, managed uptime, safety tooling, and low operational overhead matter more than benchmark scores — and switching inertia, more than any remaining capability gap, is probably the

single biggest reason token volume hasn't followed the cost curve toward open weights.

The pattern that has emerged across most serious 2026 deployments is hybrid rather than exclusive: open models absorb the high-volume, cost-sensitive, or data-sensitive middle of the workload, while a closed frontier model gets called in for the narrow slice of tasks — the hardest coding problems, the highest-stakes reasoning — that genuinely need the extra edge.

The strategic implication for builders

The build-versus-buy question is no longer open versus closed — it is which slice of a given workload actually needs frontier reasoning and which slice is commoditized enough to run on cheaper open weights. For a platform being built today, that argues for model-agnostic infrastructure from the outset, rather than betting the stack on a single provider's roadmap.

That is as much a governance and vendor-risk decision as a technical one, and it belongs in the same conversation as data residency and cost forecasting — not as an afterthought once a provider is already load-bearing.

If open models have closed most of the capability gap at a fraction of the cost, what does it say about enterprise decision-making that adoption is moving toward closed models instead of away from them?