

If an index says “Microsoft last Apr 2025” and “Amazon last Apr 2025,” it can sound as if both companies stopped building AI models. That is probably not what the index means. More likely, it has not recorded a qualifying premium release from either company since April 2025.

That distinction matters. An index is not a complete diary of every model announcement, product update, API change, or partner launch. It applies its own definition of [recent AI model releases](#) a release, and its coverage may lag behind events. To interpret the result, ask three questions: What counts as “premium”? Does the date mean announced or usable? And is the index tracking a company’s own model releases, or every model customers can access through its cloud?

Those questions are more useful than treating “no premium release” as a verdict on Microsoft’s or Amazon’s AI work. The label is a statement about a dataset and its rules—not, by itself, evidence that either company has been idle.

First, what does “no premium release” mean?

“Premium release” is not a universal industry category. An index might use it for a new, top-tier model that is available through a product or API. Another might include only major model launches with public evaluations. A third might count releases only when a model appears in a particular leaderboard or dataset.



Those rules produce different timelines. A company might release a smaller model, update an existing model, announce a research preview, or add a third-party model to its cloud without satisfying an index’s definition of a new premium release. In that case, the company did release something; the index simply does not count it in that category.

The wording also doesn’t tell you whether the tracker covers company-owned models or all models sold through a company’s services. That difference is especially important for Microsoft and Amazon. Both offer access to models from outside developers. A model becoming available in Azure or Amazon Bedrock is a cloud-platform event. It is not automatically a new Microsoft- or Amazon-developed model.

Announced is not the same as usable

Release dates are another source of confusion. A company may announce a model on one date, open a limited preview later, and make it generally available after that. The announcement date is useful for understanding when the company disclosed the model. The usable date is more useful for customers deciding when they could actually test or deploy it.

For example, Amazon announced Nova Premier in April 2025. If an index labels Amazon “last Apr 2025,” that is consistent with counting that launch as the latest qualifying Amazon model release. But the label still depends on the index’s inclusion rules and the date field it uses. It should not be read as proof that no Amazon AI product, service, or model changed afterward.

Microsoft’s April 2025 marker should be read with the same care. Check the underlying entry: which model is it referring to, what kind of availability did Microsoft report, and does the index use the announcement date or a date when users could access the model? Without those details, “Microsoft last Apr 2025” is a useful pointer, not a complete release history.

A practical tracker should preserve both dates when possible:

Milestone	What it tells you	Why it matters	Announced	When the company publicly described the release	Useful
	for tracking strategy and public commitments	Preview or limited access	When some users could start testing it		
Availability	may be restricted or change during testing	Generally available	When eligible customers could use the production offering	Usually the most relevant date for deployment decisions	Leaderboard or dataset appearance
	When the model became visible in a specific evaluation source	May come after launch and depends on participation and data updates			

What a preference leaderboard does—and does not—measure

LMarena, formerly LMSYS Chatbot Arena, is one useful place to compare models through human preference votes. In a typical blind comparison, people submit a prompt, receive answers from two unidentified models, and choose the response they prefer. The platform aggregates those preferences into rankings, with features such as style control intended to help account for differences in response style.

This is not the same as a launch benchmark. A company’s launch materials may report scores on selected coding, math, reasoning, or knowledge tests. Those tests can help answer specific questions, but they do not necessarily predict which answer people will prefer in open-ended, everyday conversations. Arena-style voting measures a different thing: how often a response wins a preference comparison under that platform’s conditions.

Neither method is a universal measure of model quality. A preference ranking can be influenced by the prompts people submit, the mix of voters, response length, formatting, and how well a model’s style matches what voters like. A benchmark score depends on the test, its scoring method, and the possibility that a model was optimized for that task. Read the method before reading the rank.

For repeatable analysis, researchers can also inspect the Hugging Face dataset `lmarena-ai/leaderboard-dataset`. It provides a way to work with leaderboard data, but it should not be treated as a complete catalog of all releases. A model needs to be represented in the relevant source, and the dataset’s coverage and update timing affect what an analyst can conclude.

Why an index can show a long gap

There are several plausible reasons Microsoft and Amazon could show an April 2025 last-release marker even while continuing to ship AI-related products or services:

- **The index has a narrow definition.** It may count only new, high-end models under each company's own name.
- **The company shipped updates, not a new model family.** Improvements to an existing model may not qualify as a fresh premium release.
- **The product is a partner model.** Cloud access to a model from OpenAI, Anthropic, or another developer does not make it a Microsoft- or Amazon-developed model.
- **The release is not yet generally available.** A preview or announcement may be excluded if the tracker requires production access.
- **The source has not caught up.** Dataset updates, model submissions, and index refreshes do not always happen on launch day.
- **The company chose a different release cadence.** A longer gap between flagship models can reflect product strategy, not a lack of engineering activity.

That last point is easy to miss. Labs do not release on a shared schedule. OpenAI, Anthropic, Google DeepMind, Microsoft, Amazon, and open-model teams make different trade-offs around training, product integration, safety work, and serving costs. Some publish frequent model variants. Others make fewer headline releases while updating products behind the scenes. Comparing cadence is useful, but it does not tell you which model is better at a particular task.

Cadence is not the same as progress

A release count is a blunt measure. Frequent releases can give customers more chances to test improvements, but they can also create churn: new model names, changing behavior, and another round of evaluation. A slower cadence can make procurement and deployment planning easier, but it can leave users waiting for capabilities they need.

The question for a team is not simply, "Who released most recently?" It is, "What changed in daily use?" Test the work your product actually depends on: tool calls, structured output, long-context retrieval, coding, latency, refusal behavior, and the quality of answers on your own prompts. A model that wins a public preference comparison may still be a poor fit for a particular workflow.

When comparing a new release with the version it replaces, keep the old version available during evaluation. Check whether the improvement is consistent across your task set, not just in a handful of impressive examples. Also look for regressions. A new model may be stronger at reasoning but worse at following a strict output schema, slower on short requests, or more expensive at the volume you expect.

How to make an upgrade decision

1. **Confirm the dates.** Record announcement, preview, and generally available dates separately. Note which one the index uses.
2. **Confirm the release type.** Is it a new model, a revised version, a product feature, or a third-party model offered through a cloud platform?
3. **Run your own evaluation.** Use a stable set of representative prompts, expected outputs, and failure cases. Keep the same prompts for the old and new versions.
4. **Measure operational changes.** Compare latency, reliability, token use, rate limits, tool behavior, and cost—not just answer quality.

5. **Check for regressions.** Include edge cases and tasks where the current model performs well. A release can improve one capability and weaken another.
6. **Make the switch reversible.** Roll out gradually, monitor production outcomes, and retain a fallback until the new version has earned confidence.

Cost belongs in that evaluation too. For example, a **secondary source reported that GPT-5.2 cost about 40% more than GPT-5.1**. Treat that as a lab claim reported by a secondary source—not as a universal price comparison. Actual cost depends on the pricing terms, token mix, usage pattern, and date of the source. Verify current rates with the provider and calculate them against your own workload before using the figure in a budget.

How to read “Microsoft last Apr 2025” and “Amazon last Apr 2025”

The most careful reading is narrow: according to the index’s current rules and recorded data, its latest qualifying premium-release entries for Microsoft and Amazon are dated April 2025. That does not establish that either company stopped releasing AI features, that customers had no new models to use through their cloud platforms, or that their systems made no progress after that month.



To verify the claim, open the index entry and check its model name, source link, date definition, and last-updated time. Then compare the entry with the company’s release notes or official announcement. If the index relies on leaderboard data, check when the model first appeared in that source and whether the ranking reflects blind-vote preference, a benchmark, or both.

This is also where Suprmind Data Lab and other teams maintaining model trackers should be precise. A good index makes its inclusion rules visible, distinguishes announcements from usable releases, and gives readers enough source information to reproduce the timeline. It should not imply that a cloud partnership is a company-built model launch, or that an absent leaderboard entry proves a model does not exist.

The short answer

The index probably says Microsoft and Amazon had “no premium release” in over a year because it has not recorded a new release that meets its specific criteria since April 2025. The date may refer to an announcement

rather than general availability, and the index may exclude product updates, previews, or partner models. Check the underlying entries before treating the label as a complete company history.

For model buyers, release cadence is a signal to investigate—not a score. Use LMArena’s blind-vote comparisons and its style-control options as one perspective, consult the leaderboard dataset for its documented coverage, and test candidate models on your own workload. The useful question is not merely whether a company has released something recently. It is whether the available model improves your daily work without introducing regressions you cannot accept.