

Tencent open-sources Hy3 preview, a 295 billion parameter Hunyuan model

Source copy kept by wheresthe.ai for the story "Tencent open-sources Hy3 preview, a 295 billion parameter Hunyuan model" (<https://wheresthe.ai/d/cn-tencent-hy3-preview-open-2026-04/>).

Source: <https://www.tencent.com/en-us/articles/2202320.html>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 53b4ce16bbca24b9f181ed720674a0211402556fa49f2e2a26f5a01177d6bfa7

Tencent today launched and open sourced the Hy3 preview model. It is a Mixture-of-Experts (MoE) model that integrates both fast and slow thinking capabilities, with a total of 295 billion parameters, 21 billion activated parameters, and supports up to a 256K token context window.

Hy3 preview is the most intelligent model in the Hy series (Hy is also known as 混元 "Hunyuan" in the China market) to date, and it has achieved substantial advancement in complex reasoning, instruction following, context learning, coding, agent capabilities, and overall inference performance.

Since February 2026, Tencent has rebuilt its pre-training and reinforcement learning infrastructure focused on making Hy models more practical for real-world use, guided by three core principles: developing well-rounded capabilities across reasoning, long-context understanding, instruction following, and tool use; prioritizing authentic evaluation beyond standard benchmarks to better reflect the model's true capabilities for continuous assessment and improvement; and integrating model and inference design with a strong business application mindset to achieve cost efficiency and bring greater value within reach.

Yao Shunyu, Chief AI Scientist at Tencent, said, "At Tencent, we are continuously expanding the scale of our pre-training and reinforcement learning efforts to push the boundaries of model intelligence. Today's Hy3 preview release allows us to gather feedback from users and the broader community, helping us further optimize the model's performance and real-world applicability ahead of its official launch. Through deep co-design with Tencent's products, we will continue to enhance real-world performance while uncovering differentiated capabilities of the model."

Hy3 preview has been integrated across Tencent's core products, including Yuanbao, ima, CodeBuddy, WorkBuddy, QQ, QQ Browser, Tencent Docs and Tencent LearnShare, with more integrations planned for products across Tencent ecosystems.

For instance, within the AI chatbot Yuanbao, Hy has been deeply co-designed with the product to significantly enhance intent understanding accuracy and text generation quality. Hy3 preview also delivers substantial performances across CodeBuddy and WorkBuddy, lowering Time To First Token by 54%, end-to-end response time by 47%, while achieving a success rate exceeding 99.99%.

In real user environments, the model has demonstrated the ability to reliably power complex agent workflows of up to 495 steps, supporting a wide range of use cases including document processing, data analysis, knowledge retrieval, and MCP (Model Context Protocol) toolchain orchestration across diverse work scenarios. In Tencent Docs' AI PPT feature, the model shows notable improvements over the previous version (Hy2), with a 20% increase in generation success rate.

Hy3 preview supports integration with popular open-source agent frameworks such as OpenClaw, OpenCode, and KiloCode, and is now available on Tencent Cloud's TokenHub large-model service platform. In addition, the Hy3 preview API has been listed on OpenRouter, with free access available for a limited period of two weeks.

Hy3 preview tops benchmarks in for overall usability and agent capabilities

Multiple evaluation results show that Hy3 preview has achieved comprehensive improvements in model capabilities.

External and in-house developed benchmarks show significantly enhanced context learning and instruction-following performance in Hy3 preview.

For STEM applications, Hy3 preview performed exceptionally on complex reasoning tasks in AI benchmarks and

real-world exams.

Hy3 preview is also proving to be one of the most practical model choices for complex agent scenarios and frameworks such as OpenClaw. Benefiting from the reconstruction of Hy's pre-training and reinforcement learning frameworks, Hy3 preview excels in agent-led tasks such as coding (performed within development environments) and search executions (retrieving, filtering and integrating information from open sources).

40% Increase in Inference Efficiency, Delivering Optimal Intelligence

Hy3 preview delivers a 40% improvement in inference efficiency while achieving optimal intelligence density at comparable cost. This is enabled by deep co-optimization between the model architecture and inference framework, along with comprehensive enhancements across the inference stack, including compute performance and quantization algorithms.

On Tencent Cloud TokenHub, Hy3 preview offers cost-effective pricing, with input costs starting at approximately USD 0.18 per million tokens, cached input costs at approximately USD 0.06 per million tokens, and output costs starting at approximately USD 0.59 per million tokens. In addition, Tencent Cloud has introduced a customized Hy3 preview Token Plan package, with personal plans starting at around USD 4.10 per month that enables use of the large model within agent development platforms and agent frameworks such as OpenClaw.

Hy3 preview is open-sourced and is now available on platforms including GitHub , Hugging Face , ModelScope , and GitCode . It supports mainstream inference frameworks such as vLLM and SGLang, enabling developers to download and deploy it directly.

Share this

Related news

See all

Corporate For business +1 more

Tencent Esports Competition Solution to Support Esports at the 20th Asian Games

September 22, 2026

Stories For business +1 more

How Weixin Mini Games Bring People Together

August 20, 2026

Corporate For business +1 more

Tencent Hy3 Now Available Globally, Extending Practical AI Across Products, Workflows and Cloud Services

August 5, 2026

Corporate For business

TenPay Global Named to CNBC's World's Top Fintech Companies 2026 List

July 23, 2026