

# Qwen3.5: One Model for Text, Images, Video, and Agent Tasks

Kabui, Charles

2026-03-06

---

[Read at ToKnow.ai](#)

---

The infographic features a dark blue background with a grid pattern. At the top left, the title 'Qwen3.5: One Model for Agent Tasks' is displayed in white and light blue. Below the title, the text 'Unified text, image, and video in a single open-weights architecture' is written in a smaller white font. Three key metrics are highlighted in separate boxes: '3B / 35B' (Active vs. total parameters, 256 MoE experts), '85.3%' (MMLU-Pro score, Frontier-competitive reasoning), and '201' (Languages and dialects, Most inclusive open model). A faint line graph is visible in the upper right corner. The 'ToKnow.ai' logo is located in the bottom right corner.

## Qwen3.5: One Model for Agent Tasks

Unified text, image, and video in a single open-weights architecture

- 3B / 35B**  
Active vs. total parameters  
256 MoE experts
- 85.3%**  
MMLU-Pro score  
Frontier-competitive reasoning
- 201**  
Languages and dialects  
Most inclusive open model

ToKnow.ai

Alibaba's Qwen team released [Qwen3.5](#), a foundation model family that merges language, vision, and agent capabilities into a single architecture through early fusion training on multimodal tokens. Previous Qwen generations split these across separate models (Qwen3 for text, Qwen3-VL for vision). The flagship [Qwen3.5-35B-A3B](#) uses a [Mixture-of-Experts](#) (MoE) layout with 256 experts, activating only 3 billion of its 35 billion total parameters per forward

pass. Its core building block is a hybrid of linear attention and gated attention called Gated Delta Networks, arranged in a 3:1 ratio with MoE layers. The model natively handles 262,144 tokens (extensible to over 1 million), covers 201 languages, and processes text, images, and video in one pass. On benchmarks: 85.3% on MMLU-Pro, 84.2% on GPQA Diamond, and 69.2% on [SWE-bench Verified](#) for real-world coding. It ships under the Apache 2.0 license and has crossed 1 million downloads on Hugging Face.

Because only 3B parameters activate at once, Qwen3.5-35B-A3B runs on consumer hardware while delivering scores that compete with models 10x its active size. For developers building AI agents, the model natively supports tool calling, MCP integration, and multi-turn agentic workflows, eliminating the need to stitch together separate language and vision models. The 201-language support makes it the most linguistically inclusive open-weights model available. [Z.ai's GLM-5](#), another recent open-source MoE model (744B total, 40B active), offers higher raw parameter counts but at significantly greater compute cost.

The pattern is clear: the gap between proprietary and open-weights models is closing, and unified multimodal architectures are replacing the patchwork of specialized models that defined the previous generation.

Read More: [RynnBrain: One Open-Source Model for Robots That See, Reason, and Act](#) — Alibaba's DAMO Academy applies a similar unified approach to embodied AI, combining perception, reasoning, and planning in a single open-source model with a 30B MoE variant.

Sources:

- [Qwen3.5 Blog Post](#)
- [Qwen3.5-35B-A3B Model Card on Hugging Face](#)
- [Qwen3.5 Collection on Hugging Face](#)
- [GLM-5 on Hugging Face](#)

---

*Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. Read more: [/terms-of-service](#)*