

RynnBrain: One Open-Source Model for Robots That See, Reason, and Act

Kabui, Charles

2026-03-04

[Read at ToKnow.ai](#)

RynnBrain: One Model for Robots That Act

See, reason, and plan in one open-source foundation model

| | | |
|---|---|--|
| 20M Training samples across 28 benchmarks | 71.0 VSI-Bench score vs 48.5 for next best open model | 3B Active params in 30B Mixture-of-Experts variant |
|---|---|--|

ToKnow.ai

Alibaba's DAMO Academy released [RynnBrain](#), an open-source foundation model built for robots and physical agents. Unlike standard vision-language models that process static images and text, RynnBrain integrates four capabilities: egocentric perception (understanding the world from the agent's viewpoint), spatiotemporal localization (tracking objects across time and space), physically grounded reasoning (connecting language to real-world coordinates), and

physics-aware planning (generating executable action sequences). The model family comes in three sizes: 2B, 8B, and a 30B Mixture-of-Experts variant that activates only 3B parameters per input. It was pretrained on roughly 20 million samples using [RynnScale](#), a load-balanced framework that doubles training throughput. Four fine-tuned variants cover navigation, manipulation planning, low-level robot control, and spatially grounded reasoning that alternates between text and coordinate predictions.

Teams building robots previously needed separate models for vision, spatial reasoning, and action planning. RynnBrain replaces that stack with one model small enough to run on-device at the 8B scale. On [RynnBrain-Bench](#), a 21-capability benchmark, the 8B model scores 71.0 on VSI-Bench spatial understanding, versus 48.5 for [MiMo-Embodied](#) and 60.3 for Qwen3-VL at the same scale. All weights, training code, fine-tuning recipes, and benchmark suite are open-sourced.

The release signals a shift in embodied AI. Until now, the best spatiotemporal models for robotics were proprietary or required assembling separate components. An open, unified backbone from perception through planning lowers the bar from “build everything” to “fine-tune one model,” accelerating prototyping across the field.

Sources:

- [RynnBrain Paper \(arXiv:2602.14979\)](#)
- [RynnBrain on Hugging Face](#)
- [RynnBrain Project Page](#)
- [RynnBrain GitHub](#)

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