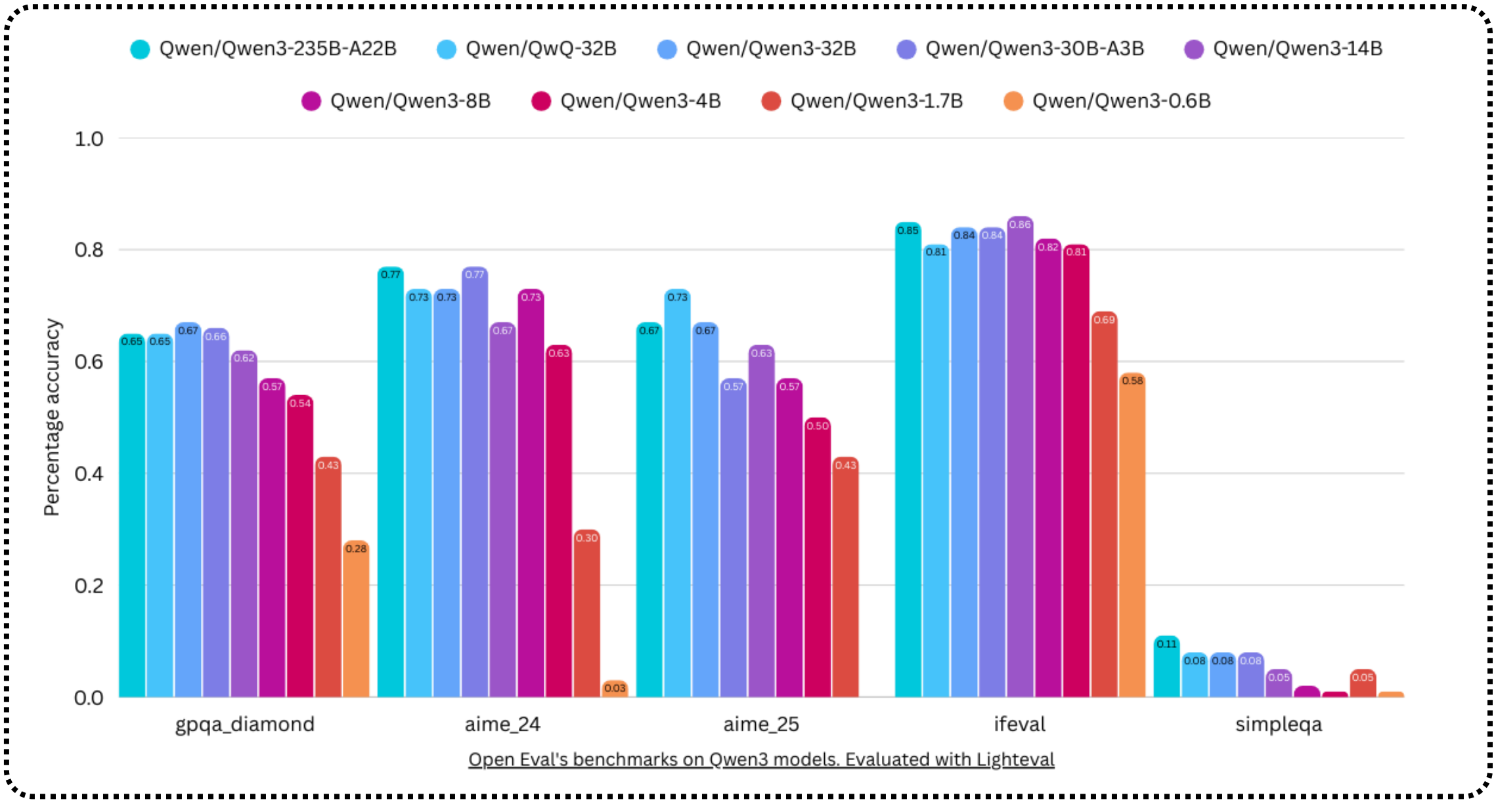
Analytics Vidhya

Qwen3 is Here

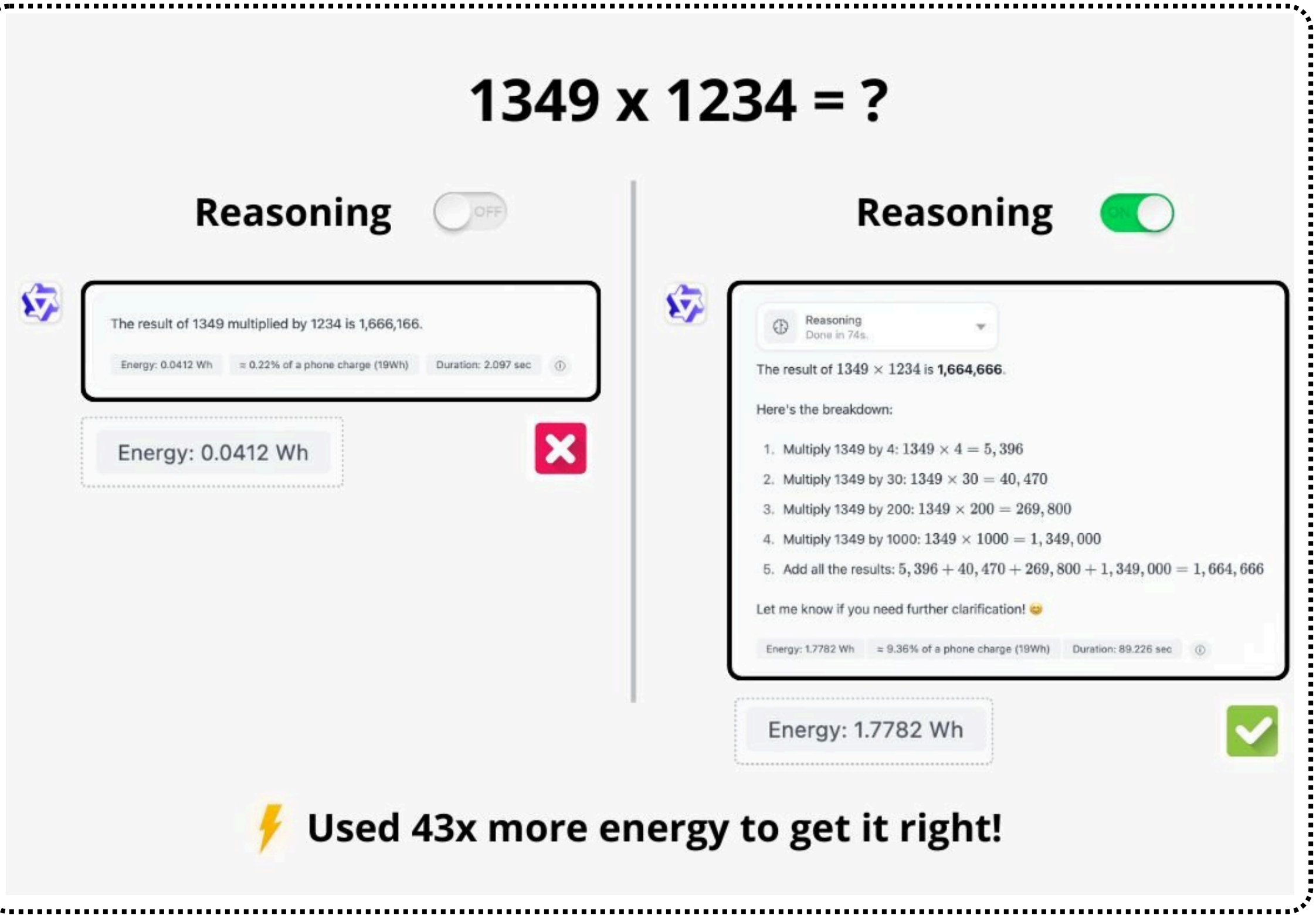
Think Deeper, Act Faster



Open Eval's benchmarks on Qwen3 models. Evaluated with Lighteval

What is Qwen3?

Qwen3 is the latest generation of large language models in Qwen series, offering a comprehensive suite of dense and **mixture-of-experts (MoE)** models. Built upon extensive training, Qwen3 delivers advancements in **reasoning**, **instruction-following**, **agent capabilities**, and **multilingual support.**



Qwen3 Models?

Qwen is open-weighting two MoE models: the large Qwen3-235B-A22B and the smaller Qwen3-30B-A3B along with **6 dense models** from Qwen3-32B down to Qwen3-0.6B.

Models	Layers	Heads (Q / KV)	Tie Embedding	Context Length
Qwen3-0.6B	28	16 / 8	Yes	32K
Qwen3-1.7B	28	16 / 8	Yes	32K
Qwen3-4B	36	32 / 8	Yes	32K
Qwen3-8B	36	32 / 8	No	128K
Qwen3-14B	40	40 / 8	No	128K
Qwen3-32B	64	64 / 8	No	128K

Models	Layers	Heads (Q / KV)	# Experts (Total / Activated)	Context Length
Qwen3-30B-A3B	48	32 / 4	128 / 8	128K
Qwen3-235B-A22B	94	64 / 4	128 / 8	128K

All models are released under the Apache 2.0 license.*

Benchmarks

	Qwen3-235B-A22B MoE	Qwen3-32B Dense	OpenAI-o1 2024-12-17	Deepseek-R1	Grok 3 Beta Think	Gemini2.5-Pro	OpenAI-o3-mini Medium
ArenaHard	95.6	93.8	92.1	93.2	-	96.4	89.0
AIME'24	85.7	81.4	74.3	79.8	83.9	92.0	79.6
AIME'25	81.5	72.9	79.2	70.0	77.3	86.7	74.8
LiveCodeBench v5, 2024.10-2025.02	70.7	65.7	63.9	64.3	70.6	70.4	66.3
CodeForces Elo Rating	2056	1977	1891	2029	-	2001	2036
Aider Pass@2	61.8	50.2	61.7	56.9	53.3	72.9	53.8
LiveBench 2024-11-25	77.1	74.9	75.7	71.6	-	82.4	70.0
BFCL v3	70.8	70.3	67.8	56.9	-	62.9	64.6
MultiIF 8 Languages	71.9	73.0	48.8	67.7	-	77.8	48.4

The main model, **Qwen3-235B-A22B**, scores well in coding, math, and general tasks, performing close to models like **DeepSeek-R1**, **o1**, **o3-mini**, **Grok-3**, and **Gemini-2.5-Pro**.

	Qwen3-30B-A3B MoE	QwQ-32B	Qwen3-4B Dense	Qwen2.5-72B-Instruct	Gemma3-27B-IT	DeepSeek-V3	GPT-4o 2024-11-20
ArenaHard	91.0	89.5	76.6	81.2	86.8	85.5	85.3
AIME'24	80.4	79.5	73.8	18.9	32.6	39.2	11.1
AIME'25	70.9	69.5	65.6	15.0	24.0	28.8	7.6
LiveCodeBench v5, 2024.10-2025.02	62.6	62.7	54.2	30.7	26.9	33.1	32.7
CodeForces Elo Rating	1974	1982	1671	859	1063	1134	864
GPQA	65.8	65.6	55.9	49.0	42.4	59.1	46.0
LiveBench 2024-11-25	74.3	72.0	63.6	51.4	49.2	60.5	52.2
BFCL v3	69.1	66.4	65.9	63.4	59.1	57.6	72.5
MultiIF 8 Languages	72.2	68.3	66.3	65.3	69.8	55.6	65.6

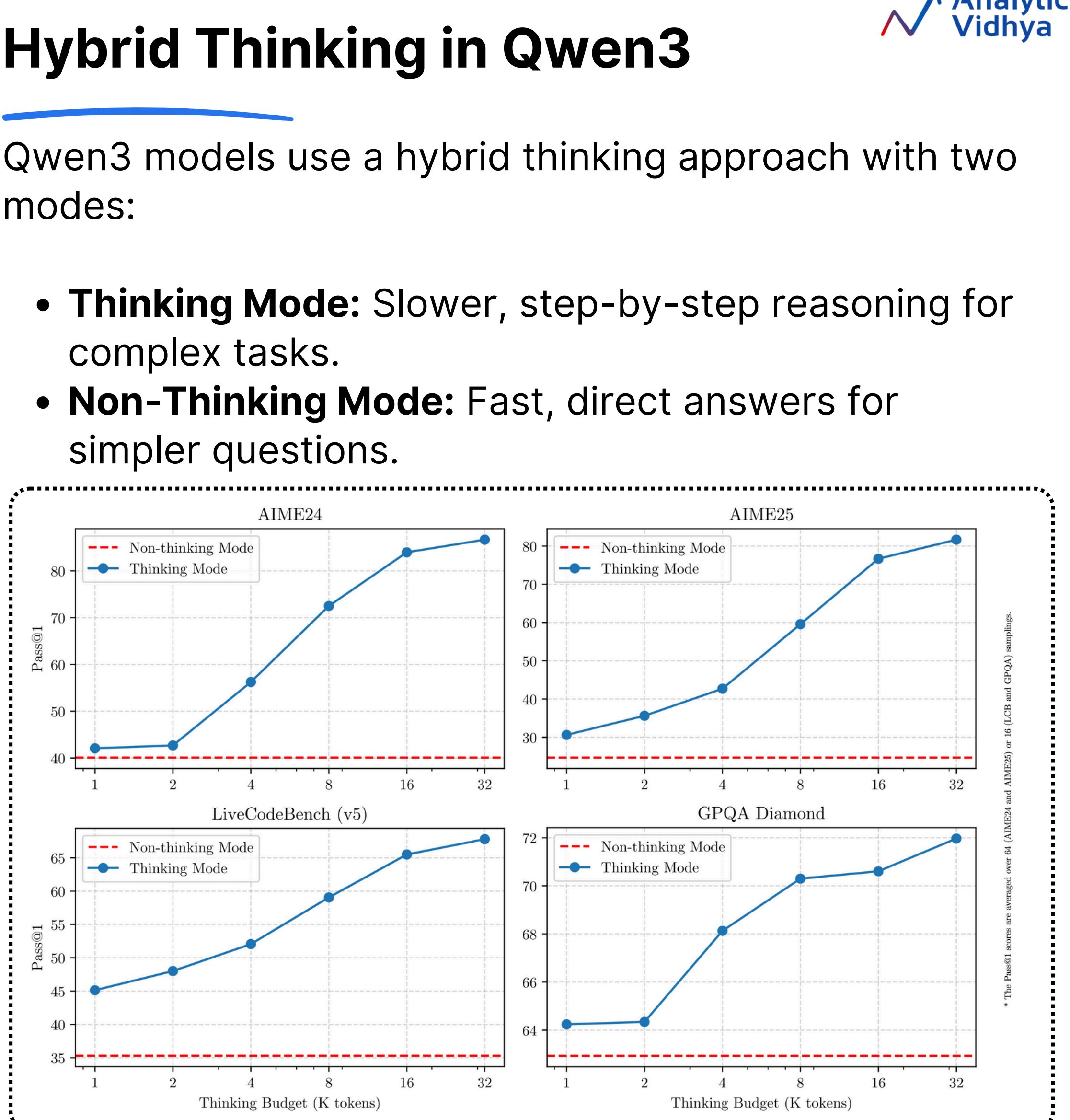
The smaller Qwen3-30B-A3B beats QwQ-32B, even with 10× fewer active parameters.

Even the tiny Qwen3-4B matches the performance of the much larger Qwen2.5-72B-Instruct.

Hybrid Thinking in Qwen3

Qwen3 models use a hybrid thinking approach with two modes:

- **Thinking Mode:** Slower, step-by-step reasoning for complex tasks.
- **Non-Thinking Mode:** Fast, direct answers for simpler questions.



The hybrid design allows users to adjust **reasoning depth**, **balancing speed** and **accuracy**. It enables stable performance, scalable thinking, and cost-effective inference by matching compute to task complexity.

Multilingual Support

Qwen3 models are supporting 119 languages and dialects. This extensive multilingual capability opens up new possibilities for international applications, enabling users worldwide to benefit from the power of these models.

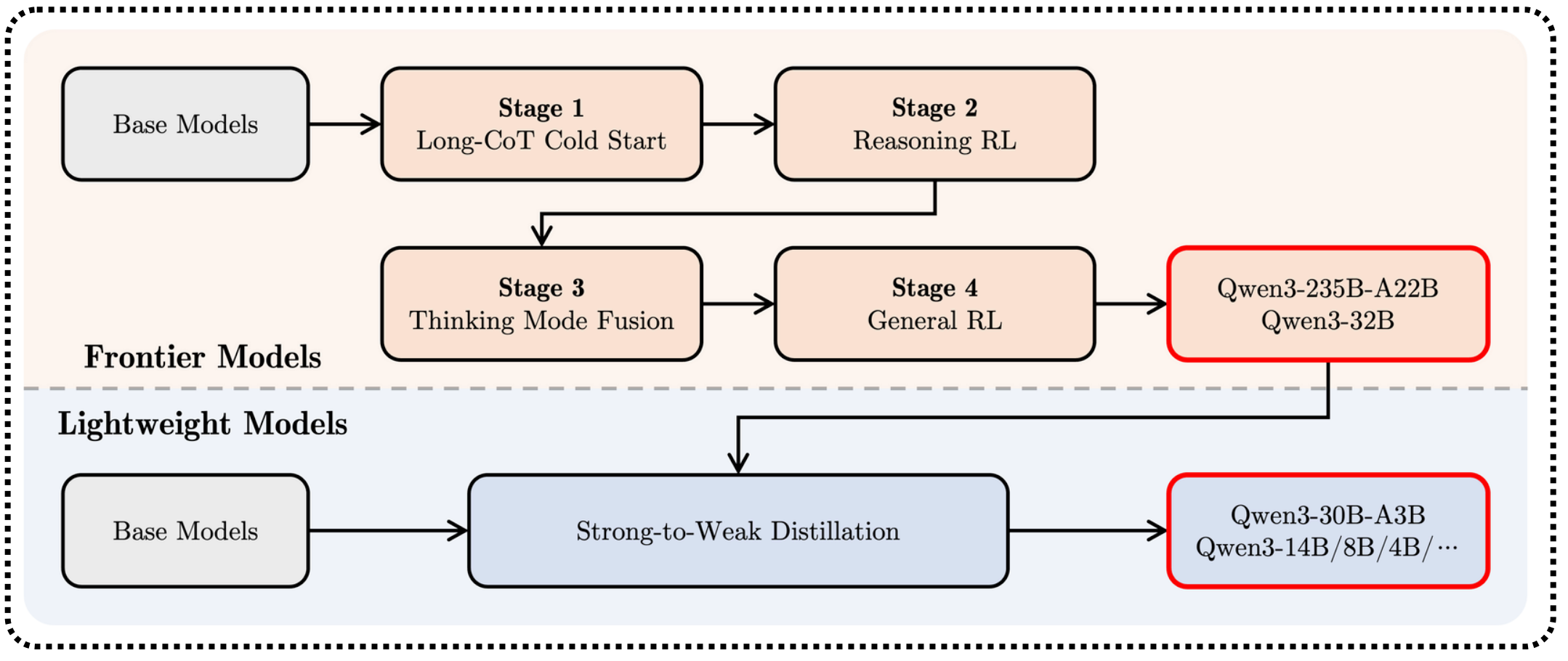
Language Family	Languages & Dialects
Indo-European	English, French, Portuguese, German, Romanian, Swedish, Danish, Bulgarian, Russian, Czech, Greek, Ukrainian, Spanish, Dutch, Slovak, Croatian, Polish, Lithuanian, Norwegian Bokmål, Norwegian Nynorsk, Persian, Slovenian, Gujarati, Latvian, Italian, Occitan, Nepali, Marathi, Belarusian, Serbian, Luxembourgish, Venetian, Assamese, Welsh, Silesian, Asturian, Chhattisgarhi, Awadhi, Maithili, Bhojpuri, Sindhi, Irish, Faroese, Hindi, Punjabi, Bengali, Oriya, Tajik, Eastern Yiddish, Lombard, Ligurian, Sicilian, Friulian, Sardinian, Galician, Catalan, Icelandic, Tosk Albanian, Limburgish, Dari, Afrikaans, Macedonian, Sinhala, Urdu, Magahi, Bosnian, Armenian
Sino-Tibetan	Chinese (Simplified Chinese, Traditional Chinese, Cantonese), Burmese
Afro-Asiatic	Arabic (Standard, Najdi, Levantine, Egyptian, Moroccan, Mesopotamian, Ta’izzi-Adeni, Tunisian), Hebrew, Maltese
Austronesian	Indonesian, Malay, Tagalog, Cebuano, Javanese, Sundanese, Minangkabau, Balinese, Banjar, Pangasinan, Iloko, Waray (Philippines)
Dravidian	Tamil, Telugu, Kannada, Malayalam
Turkic	Turkish, North Azerbaijani, Northern Uzbek, Kazakh, Bashkir, Tatar
Tai-Kadai	Thai, Lao
Uralic	Finnish, Estonian, Hungarian
Austroasiatic	Vietnamese, Khmer

Enhanced Pre and Post-Training

Pretraining

Qwen3 pretraining followed three steps: training on 30T tokens with **4K context**, then focusing on **STEM**, coding, and reasoning, and finally extending context length to **32K with long-context data**.

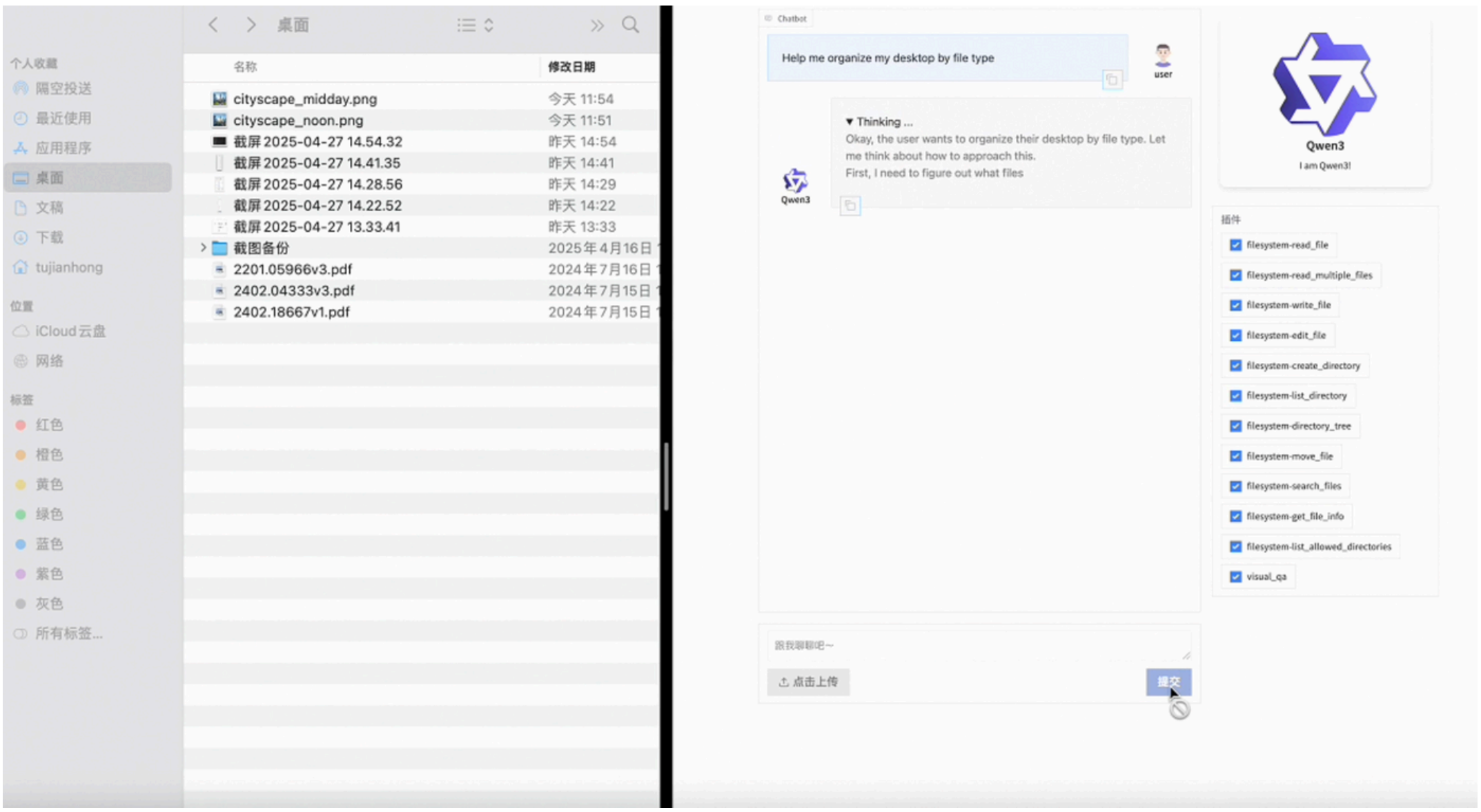
Post-training



Qwen3 models with hybrid thinking follow a 4-step reasoning process: **CoT cold start**, r**easoning-based RL**, **thinking mode fusion**, and **general RL**. Lightweight models were trained via base model distillation.

MCP & Agentic Support

The Qwen3 models have been optimized for coding and agentic capabilities, with enhanced support for MCP. The following examples demonstrate how Qwen3 reasons and interacts with its environment.



To know more, kindly visit **this article**