Alibaba Cloud

ACADEMY

Tongyi 通义

Raymond Tsang

Global Advisor in AI

v20251011

What is Tongyi?

阿里云百炼 通义大模型企业级服务平台，助力企业轻松打造最优落地效果的 AI 应用

- Alibaba Cloud provides Tongyi Qianwen (Qwen) model series to the open-source community. This series includes:

Qwen, the large language model (LLM);

Qwen-VL, the large language vision model;

Qwen-Audio, the large language audio model;

Qwen-Coder, the coding model;

Qwen-Math, the mathematical model

AI Site URL

- Tongyi 通义
  - Register by phone number / Alipay / Taobao
  - Chinese interface only
  - https://www.tongyi.com/
  - https://tongyi.aliyun.com/wanxiang/
- Qwen
  - Register by email
  - English interface only
  - https://chat.qwen.ai/
  - https://wan.video/
- Model Studio (Alibaba Cloud)
  - You can try Qwen models and easily customize and deploy them in Production.

Reason to use TongYi

Why do we use Tongyi but not others?

1. Free to use.
2. No need VPN. (For Hong Kong, Macau, China)
3. TongYi is the top model out of all the open source models. (2024 Nov,2025 May)
4. Qwen supports over 100 languages
    1. Chinese born model is an expert in Chinese Language.
    2. TongYi can understand Asian languages too. (Japanese, Korean, Thai, Indonesian, Vietnamese, Filipino)
    3. TongYi can understand European languages too. (French, Spanish, Portuguese, German, Italian, Russian, Arabic, etc.)
    4. TongYi can understand Cantonese, indeed.
5. TongYi has iOS and Android mobile app.

Reason to use TongYi

Why do we use Tongyi but not others?

1. Moreover, it is not just a chatbot. It has other functions including
   - Image generation
   - Video generation
   - Sound recording
   - Speech To Text
   - PowerPoint creation
   - Model/product replacement from image

DeepSeek vs Qwen2.5-Max

新聞首頁 / 熱門新聞

回上頁

最新全球模型榜單：阿里Qwen2.5-Max超DeepSeek V3

格 | 2025/02/05 10:28 HKT | 33 45 8 A- A+ 股價 沽空

2月5日｜2月4日凌晨，三方基準測試平台Chatbot Arena公佈了最新的大模型盲測榜單，剛剛發佈的Qwen2.5-Max超越DeepSeek V3、o1-mini和Claude-3.5-Sonnet等模型，以1332分位列全球第七名，也是非推理類的中國大模型冠軍。同時，Qwen2.5-Max在數學和編程等單項能力上排名第一，在硬提示(Hard prompts)方面排名第二。

Top of the world

2024 Nov

	Qwen2.5-72B	Qwen2-72B	Mixtral-8x22B	Llama-3-70B	LLama-3-405B
MMLU	**86.1**	84.2	77.8	79.5	85.2
MMLU-Pro	58.1	55.7	51.6	52.8	**61.6**
MMLU-redux	**83.9**	80.5	72.9	75.0	-
GPQA	**45.9**	37.4	34.3	36.3	-
BBH	**86.3**	82.4	78.9	81.0	85.9
ARC-challenge	**72.4**	68.9	70.7	68.8	-
HumanEval	59.1	**64.6**	46.3	48.2	61.0
MBPP	**84.7**	76.9	71.7	70.4	73.0
MultiPL-E	**60.5**	59.6	46.7	46.3	-
GSM8K	**91.5**	89.5	83.7	77.6	89.0
MATH	**62.1**	50.9	41.7	42.5	53.8
Multi-Exam	**78.7**	76.6	63.5	70.0	-
Multi-Understanding	**89.6**	80.7	77.7	79.9	-
Multi-Mathematics	**76.7**	76.0	62.9	67.1	-
Multi-Translation	**39.0**	37.8	23.3	38.0	-

Top of the world

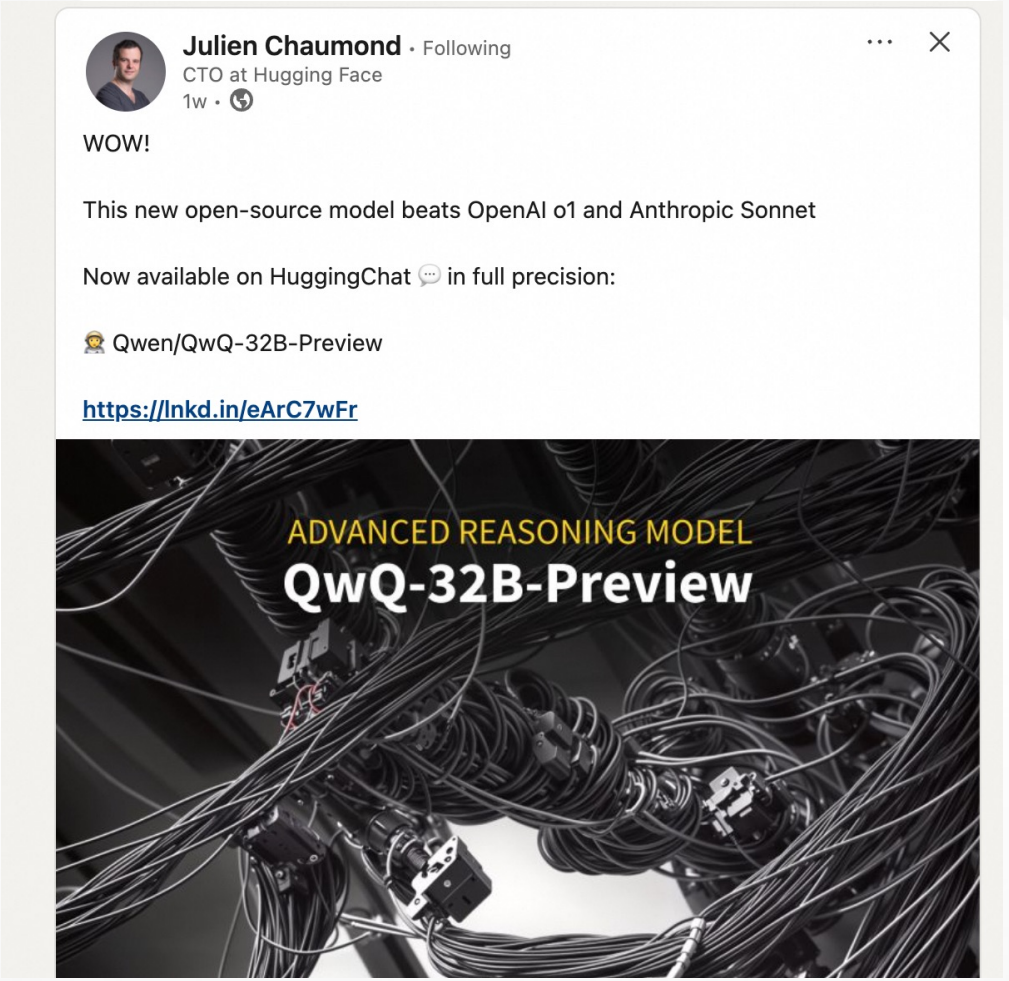
2025 May

	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

1. AIME 24/25: We sample 64 times for each query and report the average of the accuracy. AIME'25 consists of Part I and Part II, with a total of 30 questions.
2. Aider: We didn't activate the think mode of Qwen3 to balance efficiency and effectiveness.
3. BFCL: The Qwen3 models are evaluated using the FC format, while the baseline models are assessed using the highest scores obtained from either the FC or prompt formats.

Comment from Hugging Face CTO

2025 May



Alibaba Cloud

ACADEMY

THANK YOU