

2026-08-31 ??AI??????

? ???? (Executive Summary)

“ 8月26日，阿里云宣布推出新一代AI模型Qwen3.8-Flash，该模型在保持与Qwen3.8-72B同等性能的同时，推理速度提升26%，推理成本降低至原来的1/6。Qwen3.8-Flash在多个国际基准测试中表现优异，与OpenAI GPT-5.6、Anthropic Claude Haiku 4.5等顶级模型竞争。阿里云表示，Qwen3.8-Flash将广泛应用于企业级AI应用，助力客户提升效率、降低成本。 ”

#Qwen3.8-Flash #AI #阿里云 #Qwen3.8-Flash

? ?????? (Top Stories)

? ??????Qwen3.8-Flash & ??GLM-5.3-Flash 8?26???????

- 阿里云宣布推出新一代AI模型Qwen3.8-Flash，该模型在保持与Qwen3.8-72B同等性能的同时，推理速度提升26%，推理成本降低至原来的1/6。Qwen3.8-Flash在多个国际基准测试中表现优异，与OpenAI GPT-5.6、Anthropic Claude Haiku 4.5等顶级模型竞争。阿里云表示，Qwen3.8-Flash将广泛应用于企业级AI应用，助力客户提升效率、降低成本。
- AI (Z.ai) 宣布推出新一代AI模型GLM-5.3-Flash，该模型在保持与GLM-5.3-72B同等性能的同时，推理速度提升26%，推理成本降低至原来的1/6。GLM-5.3-Flash在多个国际基准测试中表现优异，与OpenAI GPT-5.6、Anthropic Claude Haiku 4.5等顶级模型竞争。Z.ai表示，GLM-5.3-Flash将广泛应用于企业级AI应用，助力客户提升效率、降低成本。
- 阿里云表示，Qwen3.8-Flash将广泛应用于企业级AI应用，助力客户提升效率、降低成本。Qwen3.8-Flash在多个国际基准测试中表现优异，与OpenAI GPT-5.6、Anthropic Claude Haiku 4.5等顶级模型竞争。
- 阿里云表示，Qwen3.8-Flash将广泛应用于企业级AI应用，助力客户提升效率、降低成本。Qwen3.8-Flash在多个国际基准测试中表现优异，与OpenAI GPT-5.6、Anthropic Claude Haiku 4.5等顶级模型竞争。

OpenAI Anthropic/Google/AWS

- 2026-08-27 OpenAI Call for Collective Action on Cyber Defense Anthropic AWS Google AI
- OpenAI
- - AI +AI B2G/B2B
 - 2023 AI AI " " "
- 8 27 Flash " " " OpenAI " "

(Track Updates)

(Models & Infra)

- NVIDIA AI Enterprise 3.0 ~8 26 AI AI (NVIDIA)
- NVIDIA " " " "
- Google DeepMind " AI " 8 (DeepMind)
- " "

(Apps & Business)

- Kimi K3 Agent 2.8 KDA Swarm Goal (Kimi)
- K3 7 16 Agent PPT/ /
- Coding Plan GLM Kimi MiMo Coding (CSDN)
- Flash Coding Plan

(Policy & Safety)

- OpenAI

- EU/ AI 8 AI (:AIApps)
 - AI

? (Radar)

	Qwen3.8-Flash Apache 2.0	125B /6B	CSDN
	8 16+ AI	Qwen3.8-Flash GLM-5.3-Flash	LLMGateway
	NVIDIA AI Enterprise 3.0	AI " " "	NVIDIA
	DeepMind AI		DeepMind
	OpenAI+Anthropic+Google + +AWS	AI " " "	
		—	—

? (Actionable Insights)

- /
 - Flash Qwen3.8-Flash ¥1/ token
 GLM-5.3-Flash Code Arena —" + "
 - Coding Agent AI
- 1. Qwen3.8-Flash-Next GLM-5.3-Flash
 - 2. Flash " "
 - 3. OpenAI
- A/ Qwen3.8-Flash GLM-5.3-Flash
 — Agent 7×24 token
 - Flash-Next Qwen4
 6B GPU RTX 4090
 125B

Al B2G

Updated 2026-08-31 06:01:29 UTC by Colin Wang