

ai.com

DeepSeek R1 破卷



时间线

11月 20 2024	DeepSeek-R1-Lite-Preview 发布 <ul style="list-style-type: none">■ AIME & MATH benchmarks 达到 v1-preview 水平■ 实时透明的思维过程■ 开源模型与 API 即将推出
12月 26 2024	DeepSeek-V3 发布 <ul style="list-style-type: none">■ 速度提升3倍 (60 tokens/second)■ 增强的能力■ 保持 API 兼容性■ 完全开源的模型和论文

时间线

1月 15 2025	DeepSeek App 发布 <ul style="list-style-type: none">■ 多平台登录支持■ 跨平台聊天记录同步■ 网页搜索和深度思考模式■ 文件上传和文本提取
1月 20 2025	DeepSeek-R1 正式发布 <ul style="list-style-type: none">■ 性能达到 OpenAI-o1 水平■ 完全开源模型和技术报告■ MIT 许可：可自由使用和商业化

国外的反应



Andrej Karpathy

OpenAI 的创始成员和研究员

这一成果展示了在数据和算法优化方面的卓越研究和工程能力，表明即使没有庞大的GPU集群，也能通过高效利用资源实现前沿模型的训练。同时，作者对技术报告的详细性表示赞赏。

国外的反应



Sam Altman
OpenAI CEO

Sam Altman (@sama) posted a tweet comparing DeepSeek's R1 model to DeepSeek's R1 model. The tweet includes two identical text snippets and ends with a reply icon and a link to the full thread.

deepseek's r1 is an impressive model, particularly around what they're able to deliver for the price.

we will obviously deliver much better models and also it's legit invigorating to have a new competitor! we will pull up some releases.

10:29 AM · Jan 28, 2025 · 14.1M Views

6.9K 10K 87K 8K

DeepSeek的R1模型性价比很高，表现令人印象深刻。

市场反应

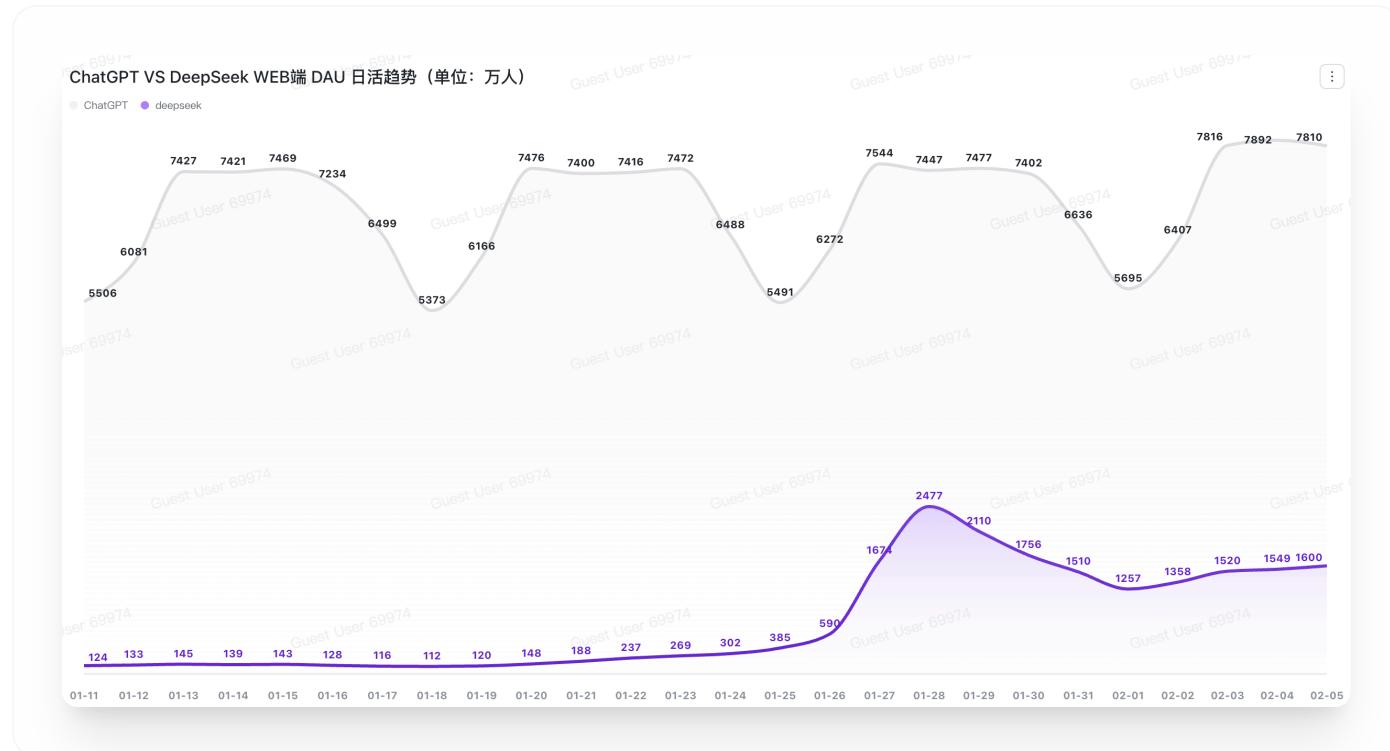
NVIDIA



中芯国际



日活跃用户



来源: <https://uniquecapital.feishu.cn/wiki/BBQMwIMfCitM1dkli9AcLWzanJb?table=blk6L0UGxdLEkLAH>

应用下载趋势

它到底是怎么发生的？

什么是推理模型

语言大模型的局限

语言大模型示例

</> 基础 Prompt

Q: 昨天是星期三，后天是星期几？

A: 星期五

♂ CoT Prompt

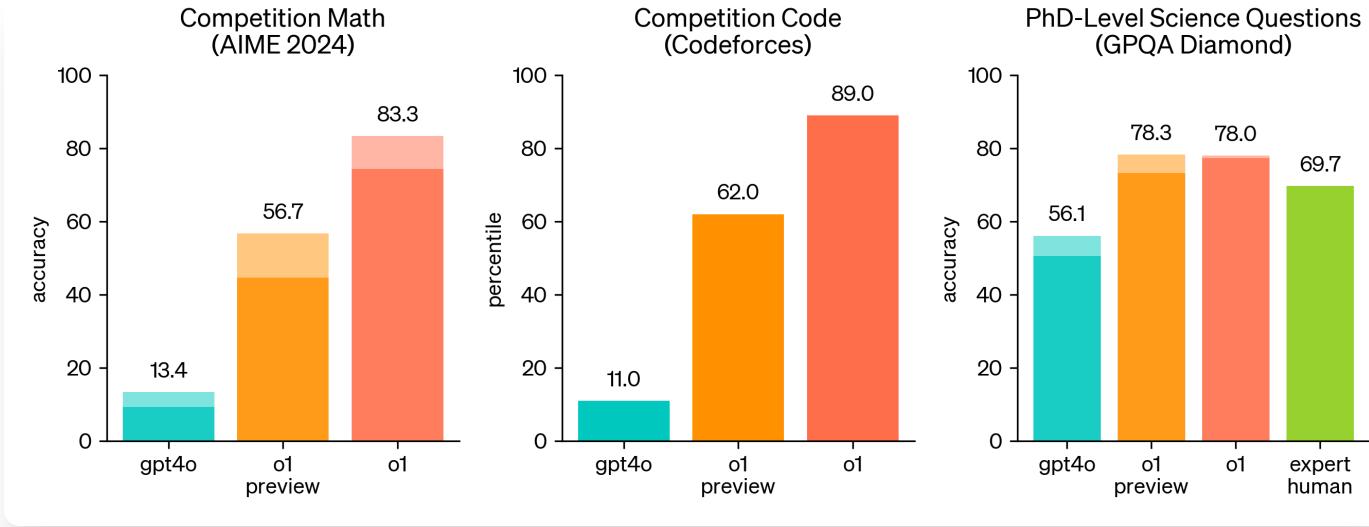
Q: 昨天是星期三，后天是星期几？如果昨天是星期三，那么今天是星期四，如果是星期四，那么明天就是星期五，后天就是星期六

A: 星期六

Chain of Thought (CoT)

第一个推理模型01

o1 benchmarks



推理模型的使用场景

谜题 & 数字证明

处理复杂的数学问题和逻辑推理

复杂决策

在多个选项中进行深入分析和选择

开放式答案

生成详细且有逻辑的解释

显式多步思考

展示清晰的推理过程

如何去复现o1

OpenAI 的最后公开的论文

过程激励模型



树搜索派系

- 多路径探索
- 动态评估



蒸馏派系

- 知识迁移
- 模型压缩

与此同时，一场精彩绝伦的探索之旅正在发生

k1.5

The Thought Process Behind Kimi k1.5

来源: https://x.com/Kimi_Moonshot/status/1882413059513471044

Kimi如何去找到灵感

Enable Self-Search

We need to enable models to search on their own!

我们需要让模型能够自主进行搜索！

Unrestricted Thinking

o1 doesn't restrict how the model thinks!

o1 并不限制模型的思考方式！

Exact Rewards

Do RL with exact rewards! Don't be constrained by Reward Models.

使用精确奖励进行强化学习！不要被奖励模型所束缚。

DeepSeek R1 的论文

DeepSeek r1-zero

强化学习算法 (Reinforcement Learning Algorithm)

GRPO

Group Relative Policy Optimization

组相对策略优化

- PPO的变体，通过组内奖励归一化来优化策略
- 提高训练效率和稳定性

训练模板(Training Template)

A conversation between User and Assistant. The user asks a question, and the Assistant solves it. The assistant first thinks about the reasoning process in the mind and then provides the user with the answer. The reasoning process and answer are enclosed within <think> </think> and <answer> </answer> tags, respectively, i.e., <think> reasoning process here </think> <answer> answer here </answer>. User: **prompt**. Assistant:

Table 1 | Template for DeepSeek-R1-Zero. **prompt** will be replaced with the specific reasoning question during training.

奖励模型 (Reward Modeling)

准确性奖励

示例

Q: $1+1=?$

正确示例 (+1分)

A: 2

错误示例 (0分)

A: 3

格式奖励

示例

Q: $1+1=?$

正确示例 (+1分)

<think> 用户问我 $1+1$ 等于几，这是一个简单的数学问题, </think>
<answer>2</answer>

错误示例 (0分)

<think></think><answer>2</answer>

模型学会了"思考"

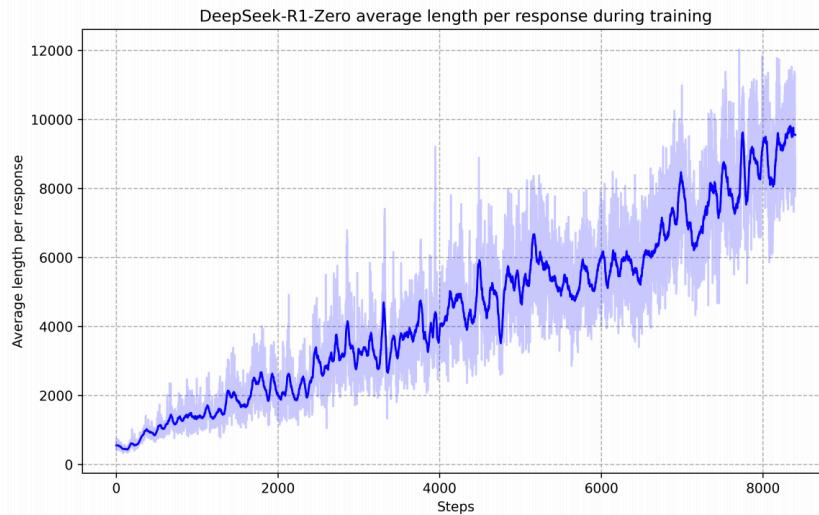


Figure 3 | The average response length of DeepSeek-R1-Zero on the training set during the RL process. DeepSeek-R1-Zero naturally learns to solve reasoning tasks with more thinking time.

r1-zero benchmarks

Model	AIME 2024		MATH-500	GPQA Diamond	LiveCode Bench	CodeForces
	pass@1	cons@64	pass@1	pass@1	pass@1	rating
OpenAI-o1-mini	63.6	80.0	90.0	60.0	53.8	1820
OpenAI-o1-0912	74.4	83.3	94.8	77.3	63.4	1843
DeepSeek-R1-Zero	71.0	86.7	95.9	73.3	50.0	1444

Table 2 | Comparison of DeepSeek-R1-Zero and OpenAI o1 models on reasoning-related benchmarks.

✓ 验证可行

r1-zero的一些问题

可读性

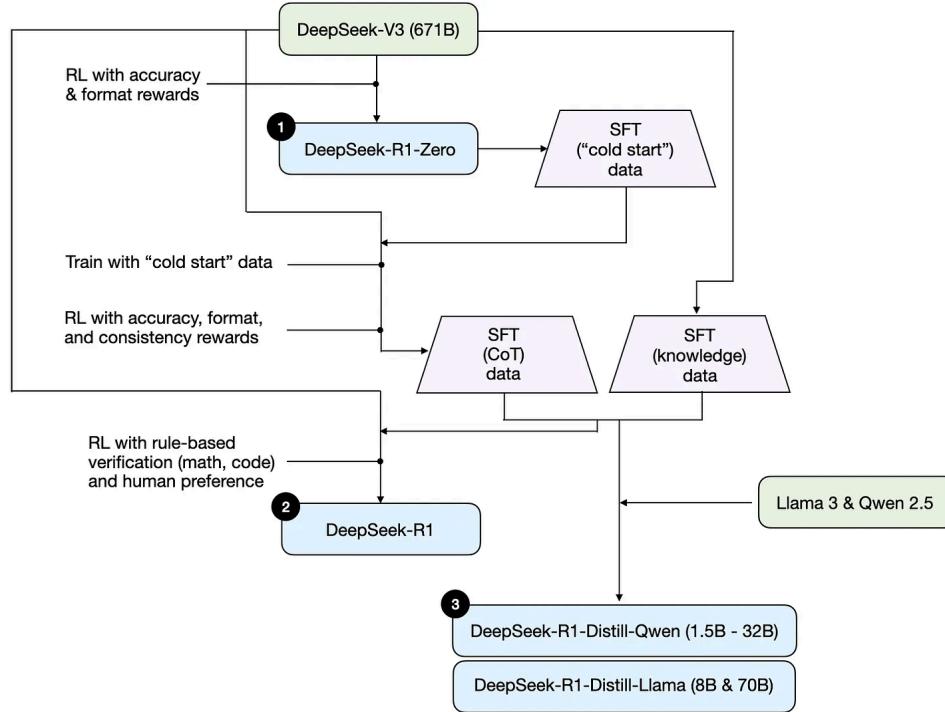
输出内容的可读性和理解性需要提升

Language Mixing

多语言混合使用导致表达不够清晰

Drawback of DeepSeek-R1-Zero Although DeepSeek-R1-Zero exhibits strong reasoning capabilities and autonomously develops unexpected and powerful reasoning behaviors, it faces several issues. For instance, DeepSeek-R1-Zero struggles with challenges like poor readability, and language mixing. To make reasoning processes more readable and share them with the open community, we explore DeepSeek-R1, a method that utilizes RL with human-friendly cold-start data.

r1的实现



r1 benchmark

Benchmark (Metric)	Claude-3.5-Sonnet-1022	GPT-4o-0513	DeepSeek-V3	OpenAI o1-mini	OpenAI o1-1217	DeepSeek R1
Architecture	-	-	MoE	-	-	MoE
# Activated Params	-	-	37B	-	-	37B
# Total Params	-	-	671B	-	-	671B
English	MMLU (Pass@1)	88.3	87.2	88.5	85.2	91.8
	MMLU-Redux (EM)	88.9	88.0	89.1	86.7	-
	MMLU-Pro (EM)	78.0	72.6	75.9	80.3	-
	DROP (3-shot F1)	88.3	83.7	91.6	83.9	90.2
	IF-Eval (Prompt Strict)	86.5	84.3	86.1	84.8	-
	GPQA Diamond (Pass@1)	65.0	49.9	59.1	60.0	75.7
	SimpleQA (Correct)	28.4	38.2	24.9	7.0	47.0
	FRAMES (Acc.)	72.5	80.5	73.3	76.9	-
	AlpacaEval2.0 (LC-winrate)	52.0	51.1	70.0	57.8	-
	ArenaHard (GPT-4-1106)	85.2	80.4	85.5	92.0	-
Code	LiveCodeBench (Pass@1-COT)	38.9	32.9	36.2	53.8	63.4
	Codeforces (Percentile)	20.3	23.6	58.7	93.4	96.6
	Codeforces (Rating)	717	759	1134	1820	2061
	SWE Verified (Resolved)	50.8	38.8	42.0	41.6	48.9
	Aider-Polyglot (Acc.)	45.3	16.0	49.6	32.9	61.7
Math	AIME 2024 (Pass@1)	16.0	9.3	39.2	63.6	79.2
	MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4
	CNMO 2024 (Pass@1)	13.1	10.8	43.2	67.6	-
Chinese	CLUEWSC (EM)	85.4	87.9	90.9	89.9	-
	C-Eval (EM)	76.7	76.0	86.5	68.9	-
	C-SimpleQA (Correct)	55.4	58.7	68.0	40.3	-

Table 4 | Comparison between DeepSeek-R1 and other representative models.

蒸馏

Model	AIME 2024		MATH-500	GPQA Diamond	LiveCode Bench	CodeForces
	pass@1	cons@64	pass@1	pass@1	pass@1	rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717
OpenAI-o1-mini	63.6	80.0	90.0	60.0	53.8	1820
QwQ-32B-Preview	50.0	60.0	90.6	54.5	41.9	1316
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481
DeepSeek-R1-Distill-Qwen-32B	72.6	83.3	94.3	62.1	57.2	1691
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205
DeepSeek-R1-Distill-Llama-70B	70.0	86.7	94.5	65.2	57.5	1633

Table 5 | Comparison of DeepSeek-R1 distilled models and other comparable models on reasoning-related benchmarks.

证明推理数据质量很高

pure rl (qwen)

Model	AIME 2024		MATH-500	GPQA Diamond	LiveCodeBench
	pass@1	cons@64	pass@1	pass@1	pass@1
QwQ-32B-Preview	50.0	60.0	90.6	54.5	41.9
DeepSeek-R1-Zero-Qwen-32B	47.0	60.0	91.6	55.0	40.2
DeepSeek-R1-Distill-Qwen-32B	72.6	83.3	94.3	62.1	57.2

Table 6 | Comparison of distilled and RL Models on Reasoning-Related Benchmarks.

deepseek v3 (不积跬步，无以至千里)

2024.2 DeepSeekMath

GRPO

2024.5 DeepSeek-V2

DeepSeekMoE

- 超大规模 21B/236B (37B/671B)
- 前置FFN
- 共享专家
- 极致的通信效率 (IB with NVLINK)

MLA (时间换空间)

- 减少KV cache
- 训练时激活内存变小
- 推理时承载更多的上下文

2024.12 DeepSeek-V3

FP8

- 减少运算量和传输

MTP

- 让模型看的更远一点，提升推理的效率

为什么要做这些创新

没卡！！！

R1 破圈的基础



强大的性能表现

达到 o1 水平的推理能力



整个业界的难题

解决推理模型的关键问题



多个方向的验证

PRM、PURE RL、Distillation



开源 & 客户端

完整的技术开放与应用支持

R1 的未来



CoT 能够激发现有模型的能力

通过思维链提升模型性能



RL 才刚刚开始

强化学习的更多可能性



基础设施终于有一个值得部署的模型

高性能与实用性的完美结合



思考时间延长

更长的推理时间，更准确的结果



Long2Short

从长文本到精炼输出的转化

Thank You