

A Survey of DeepSeek Models

唐瑞达





















目录

- DeepSeek介绍
- DeepSeek系列论文
- DeepSeek-R1

DeepSeek介绍

- DeepSeek是一家成立于2023年的中国人工智能公司，总部位于浙江省杭州市。
- 公司由幻方量化创始人梁文锋创立，并由幻方量化全资拥有。
- 从24年1月起陆续发表DeepSeek大模型相关论文。
- 在25年1月发表并开源DeepSeek-R1，打破推理模型闭源垄断。
- DeepSeek-V3和DeepSeek-R1都拥有在成本低的情况下，性能仍然顶尖的特点，和GPT等大模型性能对齐，打破了AI性能提升=算力竞赛的认知。

DeepSeek系列论文

>  DeepSeek LLM: Scaling Open-Source Language Models with Longtermism	DeepSeek-AI 等	
>  DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence	DeepSeek-AI 等	
>  DeepSeek-Coder: When the Large Language Model Meets Programming -- The Rise of Code Intelligence	Guo 等	
>  DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning	DeepSeek-AI 等	
>  DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model	DeepSeek-AI 等	
>  DeepSeek-V3 Technical Report	DeepSeek-AI 等	
>  DeepSeek-VL: Towards Real-World Vision-Language Understanding	Lu 等	
>  DeepSeek-VL2: Mixture-of-Experts Vision-Language Models for Advanced Multimodal Understanding	Wu 等	
>  DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models	Shao 等	
>  DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models	Dai 等	

- BaseModel
- Reasoning
- 多模态

DeepSeekLLM

- 24年1月发表
- 复现Llama2
- 体现DeepSeek 比较严谨的科学态度

arXiv:2401.02954v1 [cs.CL] 5 Jan 2024

DeepSeek LLM

Scaling Open-Source Language Models with Longtermism

Xiao Bi, Deli Chen, Guanting Chen, Shanhuang Chen, Damai Dai, Chengqi Deng, Honghui Ding, Kai Dong, Qiushi Du, Zhe Fu, Huazuo Gao, Kaige Gao, Wenjun Gao, Ruiqi Ge, Kang Guan, Daya Guo, Jianzhong Guo, Guangbo Hao, Zhewen Hao, Ying He, Wenjie Hu, Panpan Huang, Erhang Li, Guowei Li, Jiashi Li, Yao Li, Y.K. Li, Wenfeng Liang, Fangyun Lin, A.X. Liu, Bo Liu, Wen Liu, Xiaodong Liu, Xin Liu, Yiyuan Liu, Haoyu Lu, Shanghao Lu, Fuli Luo, Shirong Ma, Xiaotao Nie, Tian Pei, Yishi Piao, Junjie Qiu, Hui Qu, Tongzheng Ren, Zehui Ren, Chong Ruan, Zhangli Sha, Zhihong Shao, Junxiao Song, Xuecheng Su, Jingxiang Sun, Yaofeng Sun, Minghui Tang, Bingxuan Wang, Peiyi Wang, Shiyu Wang, Yaohui Wang, Yongji Wang, Tong Wu, Y. Wu, Xin Xie, Zhenda Xie, Ziwei Xie, Yiliang Xiong, Hanwei Xu, R.X. Xu, Yanhong Xu, Dejian Yang, Yuxiang You, Shuiping Yu, Xingkai Yu, B. Zhang, Haowei Zhang, Lecong Zhang, Liyue Zhang, Mingchuan Zhang, Minghua Zhang, Wentao Zhang, Yichao Zhang, Chenggang Zhao, Yao Zhao, Shangyan Zhou, Shunfeng Zhou, Qihao Zhu, Yuheng Zou *

*DeepSeek-AI

Abstract

The rapid development of open-source large language models (LLMs) has been truly remarkable. However, the scaling laws described in previous literature presents varying conclusions, which casts a dark cloud over scaling LLMs. We delve into the study of scaling laws and present our distinctive findings that facilitate the scaling of large scale models in two prevalent used open-source configurations, 7B and 67B. Guided by the scaling laws, we introduce DeepSeek LLM, a project dedicated to advancing open-source language models with a long-term perspective. To support the pre-training phase, we have developed a dataset that currently consists of 2 trillion tokens and is continuously expanding. We further conduct supervised fine-tuning (SFT) and direct preference optimization (DPO) on DeepSeek LLM Base models, resulting in the creation of DeepSeek Chat models. Our evaluation results demonstrate that DeepSeek LLM 67B surpasses LLaMA-2 70B across a range of benchmarks, especially in the domains of code, mathematics, and reasoning. Furthermore, open-ended evaluations reveal that our DeepSeek LLM 67B Chat exhibits superior performance compared to GPT-3.5.

DeepSeekMoE

- 24年1月发表
- 提出混合专家模型MoE



DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models

Damai Dai^{*1,2}, Chengqi Deng¹, Chenggang Zhao^{*1,3}, R.X. Xu¹, Huazuo Gao¹, Deli Chen¹, Jiashi Li¹, Wangding Zeng¹, Xingkai Yu^{*1,4}, Y. Wu¹, Zhenda Xie¹, Y.K. Li¹, Panpan Huang¹, Fuli Luo¹, Chong Ruan¹, Zhifang Sui², Wenfeng Liang¹

¹DeepSeek-AI

²National Key Laboratory for Multimedia Information Processing, Peking University

³Institute for Interdisciplinary Information Sciences, Tsinghua University

⁴National Key Laboratory for Novel Software Technology, Nanjing University

{daidamai, szf}@pku.edu.cn, {wenfeng.liang}@deepseek.com

<https://github.com/deepseek-ai/DeepSeek-MoE>

Abstract

In the era of large language models, Mixture-of-Experts (MoE) is a promising architecture for managing computational costs when scaling up model parameters. However, conventional MoE architectures like GShard, which activate the top- K out of N experts, face challenges in ensuring expert specialization, i.e. each expert acquires non-overlapping and focused knowledge. In response, we propose the DeepSeekMoE architecture towards ultimate expert specialization. It involves two principal strategies: (1) finely segmenting the experts into mN ones and activating mK from them, allowing for a more flexible combination of activated experts; (2) isolating K_s experts as shared ones, aiming at capturing common knowledge and mitigating redundancy in routed experts. Starting from a modest scale with 2B parameters, we demonstrate that DeepSeekMoE 2B achieves comparable performance with GShard 2.9B, which has 1.5 \times expert parameters and computation. In addition, DeepSeekMoE 2B nearly approaches the performance of its dense counterpart with the same number of total parameters, which set the upper bound of MoE models. Subsequently, we scale up DeepSeekMoE to 16B parameters and show that it achieves comparable performance with LLaMA2 7B, with only about 40% of computations. Further, our preliminary efforts to scale up DeepSeekMoE to 145B parameters consistently validate its substantial advantages over the GShard architecture, and show its performance comparable with DeepSeek 67B, using only 28.5% (maybe even 18.2%) of computations.

DeepSeek-V2

- 24年1月发表
- 提出多头注意力机制MLP
- 把混合专家的数量大幅提高到160个
- 236B参数



DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models

Damai Dai^{*1,2}, Chengqi Deng¹, Chenggang Zhao^{*1,3}, R.X. Xu¹, Huazuo Gao¹, Deli Chen¹, Jiashi Li¹, Wangding Zeng¹, Xingkai Yu^{*1,4}, Y. Wu¹, Zhenda Xie¹, Y.K. Li¹, Panpan Huang¹, Fuli Luo¹, Chong Ruan¹, Zhifang Sui², Wenfeng Liang¹

¹DeepSeek-AI

²National Key Laboratory for Multimedia Information Processing, Peking University

³Institute for Interdisciplinary Information Sciences, Tsinghua University

⁴National Key Laboratory for Novel Software Technology, Nanjing University

{daidamai, szf}@pku.edu.cn, {wenfeng.liang}@deepseek.com

<https://github.com/deepseek-ai/DeepSeek-MoE>

Abstract

In the era of large language models, Mixture-of-Experts (MoE) is a promising architecture for managing computational costs when scaling up model parameters. However, conventional MoE architectures like GShard, which activate the top- K out of N experts, face challenges in ensuring expert specialization, i.e. each expert acquires non-overlapping and focused knowledge. In response, we propose the DeepSeekMoE architecture towards ultimate expert specialization. It involves two principal strategies: (1) finely segmenting the experts into mN ones and activating mK from them, allowing for a more flexible combination of activated experts; (2) isolating K_s experts as shared ones, aiming at capturing common knowledge and mitigating redundancy in routed experts. Starting from a modest scale with 2B parameters, we demonstrate that DeepSeekMoE 2B achieves comparable performance with GShard 2.9B, which has 1.5 \times expert parameters and computation. In addition, DeepSeekMoE 2B nearly approaches the performance of its dense counterpart with the same number of total parameters, which set the upper bound of MoE models. Subsequently, we scale up DeepSeekMoE to 16B parameters and show that it achieves comparable performance with LLaMA2 7B, with only about 40% of computations. Further, our preliminary efforts to scale up DeepSeekMoE to 145B parameters consistently validate its substantial advantages over the GShard architecture, and show its performance comparable with DeepSeek 67B, using only 28.5% (maybe even 18.2%) of computations.

DeepSeek-V3

- 24年12月发表
- 训练出671B的大模型
- 仅仅使用了2000张H800
- 做负载均衡时使用了新方法
multiple token prediction
(MTP)



DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models

Damai Dai^{*1,2}, Chengqi Deng¹, Chenggang Zhao^{*1,3}, R.X. Xu¹, Huazuo Gao¹, Deli Chen¹, Jiashi Li¹, Wangding Zeng¹, Xingkai Yu^{*1,4}, Y. Wu¹, Zhenda Xie¹, Y.K. Li¹, Panpan Huang¹, Fuli Luo¹, Chong Ruan¹, Zhifang Sui², Wenfeng Liang¹

¹DeepSeek-AI

²National Key Laboratory for Multimedia Information Processing, Peking University

³Institute for Interdisciplinary Information Sciences, Tsinghua University

⁴National Key Laboratory for Novel Software Technology, Nanjing University

{daidamai, szf}@pku.edu.cn, {wenfeng.liang}@deepseek.com

<https://github.com/deepseek-ai/DeepSeek-MoE>

Abstract

In the era of large language models, Mixture-of-Experts (MoE) is a promising architecture for managing computational costs when scaling up model parameters. However, conventional MoE architectures like GShard, which activate the top- K out of N experts, face challenges in ensuring expert specialization, i.e. each expert acquires non-overlapping and focused knowledge. In response, we propose the DeepSeekMoE architecture towards ultimate expert specialization. It involves two principal strategies: (1) finely segmenting the experts into mN ones and activating mK from them, allowing for a more flexible combination of activated experts; (2) isolating K_s experts as shared ones, aiming at capturing common knowledge and mitigating redundancy in routed experts. Starting from a modest scale with 2B parameters, we demonstrate that DeepSeekMoE 2B achieves comparable performance with GShard 2.9B, which has 1.5 \times expert parameters and computation. In addition, DeepSeekMoE 2B nearly approaches the performance of its dense counterpart with the same number of total parameters, which set the upper bound of MoE models. Subsequently, we scale up DeepSeekMoE to 16B parameters and show that it achieves comparable performance with LLaMA2 7B, with only about 40% of computations. Further, our preliminary efforts to scale up DeepSeekMoE to 145B parameters consistently validate its substantial advantages over the GShard architecture, and show its performance comparable with DeepSeek 67B, using only 28.5% (maybe even 18.2%) of computations.

DeepSeek-Coder

- 24年1月、7月发表
- 为checkpoint过渡版本

4196v2 [cs.SE] 26 Jan 2024

[cs.SE] 17 Jun 2024



DeepSeek-Coder: When the Large Language Model Meets Programming - The Rise of Code Intelligence



DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence

Qihao Zhu*, Daya Guo*, Zhihong Shao*, Dejian Yang*, Peiyi Wang, Runxin Xu, Y. Wu
Yukun Li, Huazuo Gao, Shirong Ma, Wangding Zeng, Xiao Bi, Zihui Gu, Hanwei Xu, Damai Dai
Kai Dong, Liyue Zhang, Yishi Piao, Zhibin Gou, Zhenda Xie, Zhewen Hao, Bingxuan Wang
Junxiao Song, Deli Chen, Xin Xie, Kang Guan, Yuxiang You, Aixin Liu, Qiusi Du, Wenjun Gao
Xuan Lu, Qinyu Chen, Yaohui Wang, Chengqi Deng, Jiashi Li, Chenggang Zhao
Chong Ruan, Fuli Luo, Wenfeng Liang

DeepSeek-AI

<https://github.com/deepseek-ai/DeepSeek-Coder-V2>

Abstract

We present DeepSeek-Coder-V2, an open-source Mixture-of-Experts (MoE) code language model that achieves performance comparable to GPT4-Turbo in code-specific tasks. Specifically, DeepSeek-Coder-V2 is further pre-trained from an intermediate checkpoint of DeepSeek-V2 with additional 6 trillion tokens. Through this continued pre-training, DeepSeek-Coder-V2 substantially enhances the coding and mathematical reasoning capabilities of DeepSeek-V2, while maintaining comparable performance in general language tasks. Compared to DeepSeek-Coder-33B, DeepSeek-Coder-V2 demonstrates significant advancements in various aspects of code-related tasks, as well as reasoning and general capabilities. Additionally, DeepSeek-Coder

DeepSeek-VL

- 24年3月、12月发表
- 为开源视觉语言的多模态模型

arXiv:2403.05525v2 [cs.AI] 11 Mar 2024



DeepSeek-VL: Towards Real-World Vision-Language Understanding

Haoyu Lu^{*1†}, Wen I
Tongzheng Ren^{1†}, Zhuo

We present DeepSeek-Vision and language u
dimensions:

- **Data Constructio**
real-world scenarios
content (expert knowl
contexts. Further, we
instruction-tuning dat
the model's user expe

- **Model Architect**
DeepSeek-VL incorpc
images (1024 x 1024) v
tional overhead. This
detailed information a

- **Training Strateg**
possess strong langu
pretraining, we inves
from the beginning an
and language modaliti
a balanced integratio
The DeepSeek-VL far
vision-language chat
performance across a
maintaining robust p
and 7B models public

arXiv:2412.10302v1 [cs.CV] 13 Dec 2024



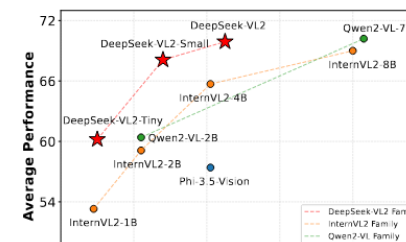
DeepSeek-VL2: Mixture-of-Experts Vision-Language Models for Advanced Multimodal Understanding

Zhiyu Wu*, Xiaokang Chen*, Zizheng Pan*, Xingchao Liu*, Wen Liu^{*†}, Damai Dai, Huazuo Gao,
Yiyang Ma, Chengyue Wu, Bingxuan Wang, Zhenda Xie, Yu Wu, Kai Hu, Jiawei Wang, Yaofeng Sun,
Yukun Li, Yishi Piao, Kang Guan, Aixin Liu, Xin Xie, Yuxiang You, Kai Dong, Xingkai Yu, Haowei Zhang,
Liang Zhao, Yisong Wang, Chong Ruan[‡]

DeepSeek-AI

Abstract

We present DeepSeek-VL2, an advanced series of large Mixture-of-Experts (MoE) Vision-Language Models that significantly improves upon its predecessor, DeepSeek-VL, through two key major upgrades. For the vision component, we incorporate a dynamic tiling vision encoding strategy designed for processing high-resolution images with different aspect ratios. For the language component, we leverage DeepSeekMoE models with the Multi-head Latent Attention mechanism, which compresses Key-Value cache into latent vectors, to enable efficient inference and high throughput. Trained on an improved vision-language dataset, DeepSeek-VL2 demonstrates superior capabilities across various tasks, including but not limited to visual question answering, optical character recognition, document/table/chart understanding, and visual grounding. Our model series is composed of three variants: DeepSeek-VL2-Tiny, DeepSeek-VL2-Small and DeepSeek-VL2, with 1.0B, 2.8B and 4.5B activated parameters respectively. DeepSeek-VL2 achieves competitive or state-of-the-art performance with similar or fewer activated parameters compared to existing open-source dense and MoE-based models. Codes and pre-trained models are publicly accessible at <https://github.com/deepseek-ai/DeepSeek-VL2>.



DeepSeekMath

- 24年2月发表
- 提出了GRPO强化学习算法

arXiv:2403.05525v2 [cs.AI] 11 Mar 2024



DeepSeek-VL: Towards Real-World Vision-Language Understanding

Haoyu Lu^{*1†}, Wen I
Tongzheng Ren^{1†}, Zhuo

We present DeepSeek-Vision and language u
dimensions:

- **Data Constructio**
real-world scenarios
content (expert knowl
contexts. Further, we
instruction-tuning dat
the model's user expe
- **Model Architect**
DeepSeek-VL incorpc
images (1024 x 1024) v
tional overhead. This
detailed information a

• **Training Strateg**
possess strong langu
pretraining, we inves
from the beginning an
and language modaliti
a balanced integratio
The DeepSeek-VL far
vision-language chat
performance across a
maintaining robust p
and 7B models public

arXiv:2412.10302v1 [cs.CV] 13 Dec 2024



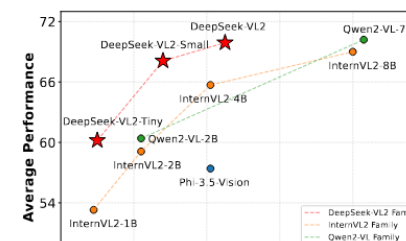
DeepSeek-VL2: Mixture-of-Experts Vision-Language Models for Advanced Multimodal Understanding

Zhiyu Wu*, Xiaokang Chen*, Zizheng Pan*, Xingchao Liu*, Wen Liu^{*†}, Damai Dai, Huazuo Gao,
Yiyang Ma, Chengyue Wu, Bingxuan Wang, Zhenda Xie, Yu Wu, Kai Hu, Jiawei Wang, Yaofeng Sun,
Yukun Li, Yishi Piao, Kang Guan, Aixin Liu, Xin Xie, Yuxiang You, Kai Dong, Xingkai Yu, Haowei Zhang,
Liang Zhao, Yisong Wang, Chong Ruan[‡]

DeepSeek-AI

Abstract

We present DeepSeek-VL2, an advanced series of large Mixture-of-Experts (MoE) Vision-Language Models that significantly improves upon its predecessor, DeepSeek-VL, through two key major upgrades. For the vision component, we incorporate a dynamic tiling vision encoding strategy designed for processing high-resolution images with different aspect ratios. For the language component, we leverage DeepSeekMoE models with the Multi-head Latent Attention mechanism, which compresses Key-Value cache into latent vectors, to enable efficient inference and high throughput. Trained on an improved vision-language dataset, DeepSeek-VL2 demonstrates superior capabilities across various tasks, including but not limited to visual question answering, optical character recognition, document/table/chart understanding, and visual grounding. Our model series is composed of three variants: DeepSeek-VL2-Tiny, DeepSeek-VL2-Small and DeepSeek-VL2, with 1.0B, 2.8B and 4.5B activated parameters respectively. DeepSeek-VL2 achieves competitive or state-of-the-art performance with similar or fewer activated parameters compared to existing open-source dense and MoE-based models. Codes and pre-trained models are publicly accessible at <https://github.com/deepseek-ai/DeepSeek-VL2>.



DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

- DeepSeek发布的第一代推理模型
- 性能对齐甚至超越GPT-o1
- 模型开源
- 使用DeepSeek自研的GRPO方法进行模型训练，只用纯强化学习来提升语言模型的推理能力
- 使用多阶段训练方式进行模型训练
- 提出有效的模型蒸馏方法

摘要

- DeepSeek推出了第一代推理模型，称为R1，其中有DeepSeek-R1-Zero 和 DeepSeek-R1。DeepSeek-R1-Zero没有使用SFT，只通过RL就展现了卓越的推理能力。它在各个数据集中分数很高，但是存在可读性差和语言混杂等问题。
- 所以这里推出DeepSeek-R1解决了这些问题，它在强化学习之前融入了多阶段训练和冷启动数据。DeepSeek-R1的推理能力可以和 OpenAI 的 o1-1217 相媲美。
- DeepSeek这里开源了DeepSeek-R1-Zero 和 DeepSeek-R1，并且使用DeepSeek-R1来得到蒸馏的SFT数据，用于Qwen和Llama系列的小模型指令微调，得到了众多蒸馏模型，在推理能力上有巨大提升，参数版本为1.5B、7B、8B、14B、32B、70B。

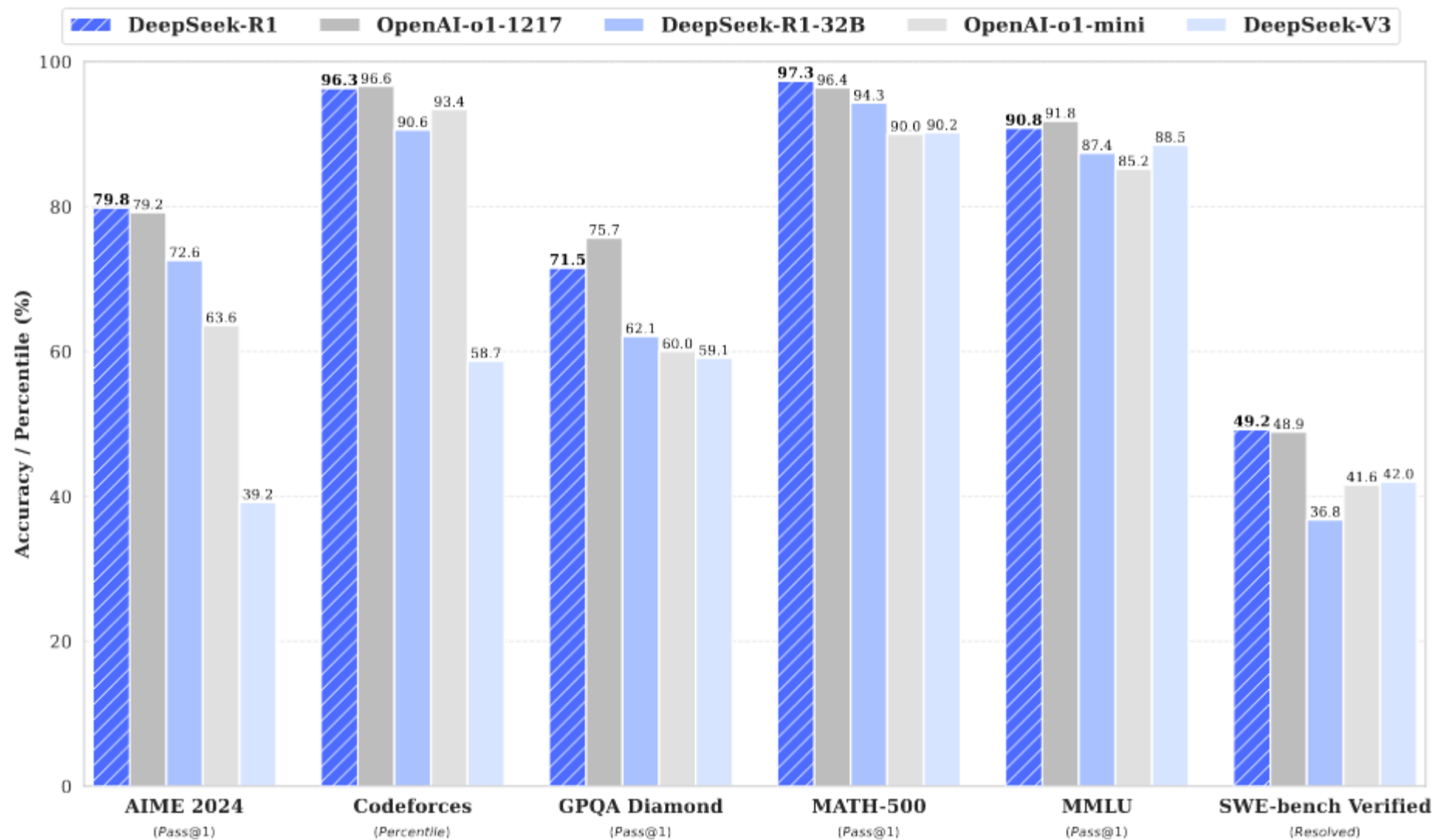


Figure 1 | Benchmark performance of DeepSeek-R1.

DeepSeek R1论文脉络梳理

实验一：训练R1 Zero

回顾&介绍GRPO
强化学习算法

训练DeepSeek R1 Zero模型

实验二：训练R1

结合R1-Zero训练
暴露的问题

训练DeepSeek R1模型

实验三：进行R1模型蒸馏

DeepSeek R1

DeepSeek R1 Distill Qwen
DeepSeek R1 Distill Llama

Contents

1 Introduction

摘要与介绍

1.1 Contributions

1.2 Summary of Evaluation Results

2 Approach

R1-Zero训练过程

2.1 Overview

2.2 DeepSeek-R1-Zero: Reinforcement Learning on the Base Model

2.2.1 Reinforcement Learning Algorithm

2.2.2 Reward Modeling

2.2.3 Training Template

2.2.4 Performance, Self-evolution Process and Aha Moment of DeepSeek-R1-Zero

2.3 DeepSeek-R1: Reinforcement Learning with Cold Start

2.3.1 Cold Start

2.3.2 Reasoning-oriented Reinforcement Learning

2.3.3 Rejection Sampling and Supervised Fine-Tuning

2.3.4 Reinforcement Learning for all Scenarios

2.4 Distillation: Empower Small Models with Reasoning Capability

R1训练过程

3 Experiment

模型蒸馏实验

3.1 DeepSeek-R1 Evaluation

3.2 Distilled Model Evaluation

4 Discussion

其他讨论

4.1 Distillation v.s. Reinforcement Learning

4.2 Unsuccessful Attempts

1. 训练DeepSeek R1 Zero

第一阶段：训练DeepSeek R1 Zero



GRPO (Group Relative Policy Optimization) : deepseek系列一贯的强化学习算法，该方法不需要类似PPO里面的评论家模型（与策略模型大小相当），极大节省训练成本。具体来说，对于每个问题 q ，GRPO从旧的策略模型（类似PPO）中采样一组输出 $\{o_1, o_2, \dots, o_G\}$ ，然后通过最大化以下目标优化策略模型

$$\mathcal{J}_{GRPO}(\theta) = \mathbb{E}[q \sim P(Q), \{o_i\}_{i=1}^G \sim \pi_{\theta_{old}}(O|q)]$$

$$\frac{1}{G} \sum_{i=1}^G \left(\min \left(\frac{\pi_{\theta}(o_i|q)}{\pi_{\theta_{old}}(o_i|q)} A_i, \text{clip} \left(\frac{\pi_{\theta}(o_i|q)}{\pi_{\theta_{old}}(o_i|q)}, 1 - \varepsilon, 1 + \varepsilon \right) A_i \right) - \beta \text{D}_{KL}(\pi_{\theta} || \pi_{ref}) \right),$$

(1)

$$\text{D}_{KL}(\pi_{\theta} || \pi_{ref}) = \frac{\pi_{ref}(o_i|q)}{\pi_{\theta}(o_i|q)} - \log \frac{\pi_{ref}(o_i|q)}{\pi_{\theta}(o_i|q)} - 1, \quad (2)$$

$$A_i = \frac{r_i - \text{mean}(\{r_1, r_2, \dots, r_G\})}{\text{std}(\{r_1, r_2, \dots, r_G\})}. \quad (3)$$

这里仅简要说明，详细GRPO讲解可查阅其他资料或等待后续制作。

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.

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.

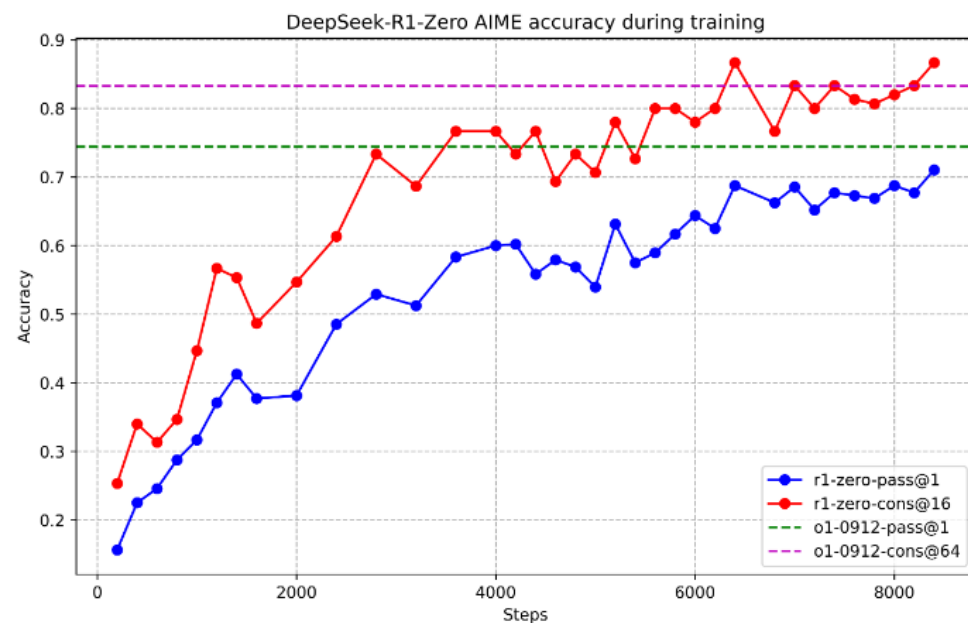


Figure 2 | AIME accuracy of DeepSeek-R1-Zero during training. For each question, we sample 16 responses and calculate the overall average accuracy to ensure a stable evaluation.

Question: If $a > 1$, then the sum of the real solutions of $\sqrt{a - \sqrt{a + x}} = x$ is equal to

Response: <think>

To solve the equation $\sqrt{a - \sqrt{a + x}} = x$, let's start by squaring both \dots

$$\left(\sqrt{a - \sqrt{a + x}}\right)^2 = x^2 \implies a - \sqrt{a + x} = x^2.$$

Rearrange to isolate the inner square root term:

$$(a - x^2)^2 = a + x \implies a^2 - 2ax^2 + (x^2)^2 = a + x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$$

\dots

Wait, wait. Wait. That's an aha moment I can flag here.

Let's reevaluate this step-by-step to identify if the correct sum can be \dots

We started with the equation:

$$\sqrt{a - \sqrt{a + x}} = x$$

First, let's square both sides:

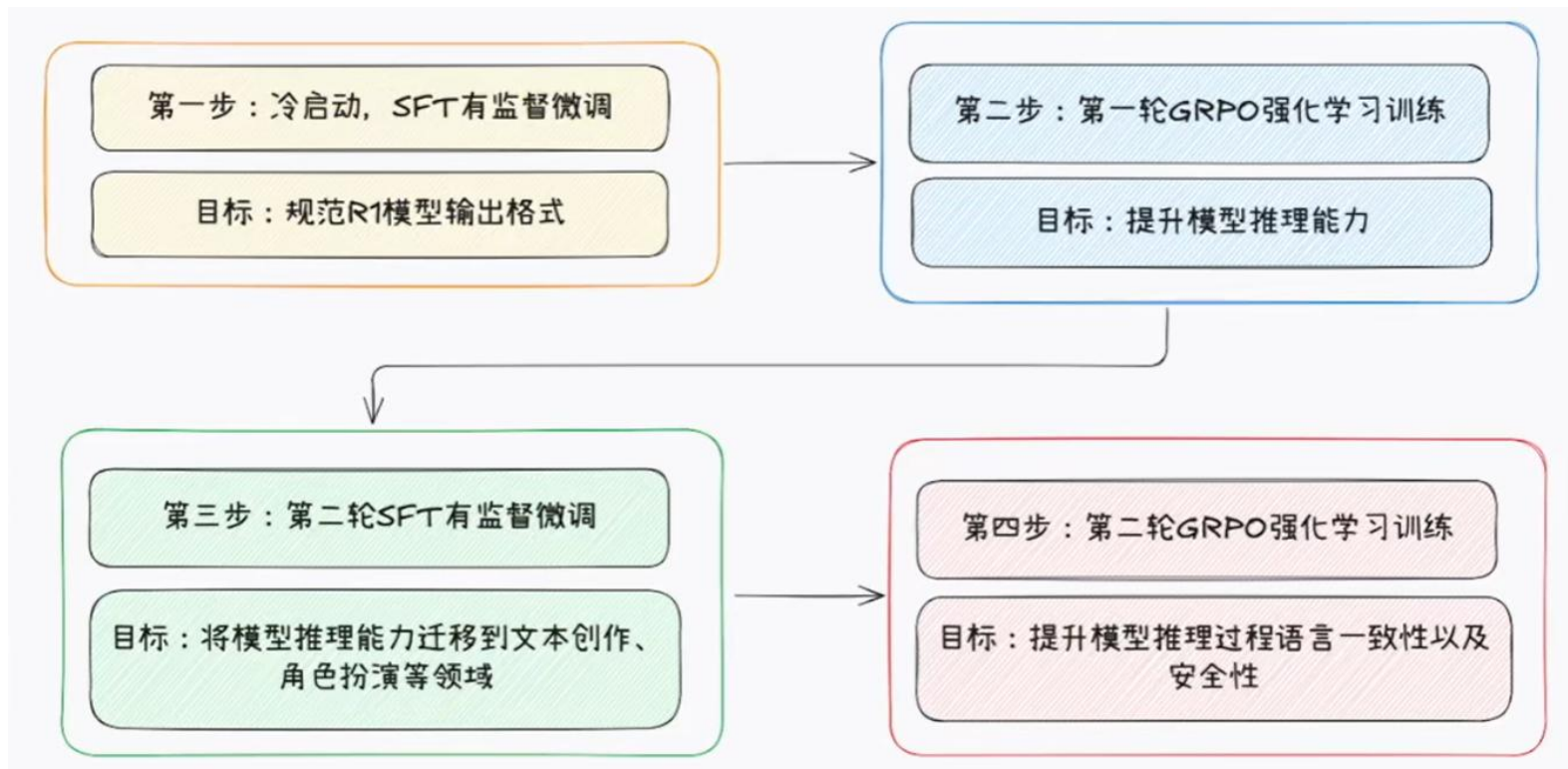
$$a - \sqrt{a + x} = x^2 \implies \sqrt{a + x} = a - x^2$$

Next, I could square both sides again, treating the equation: \dots

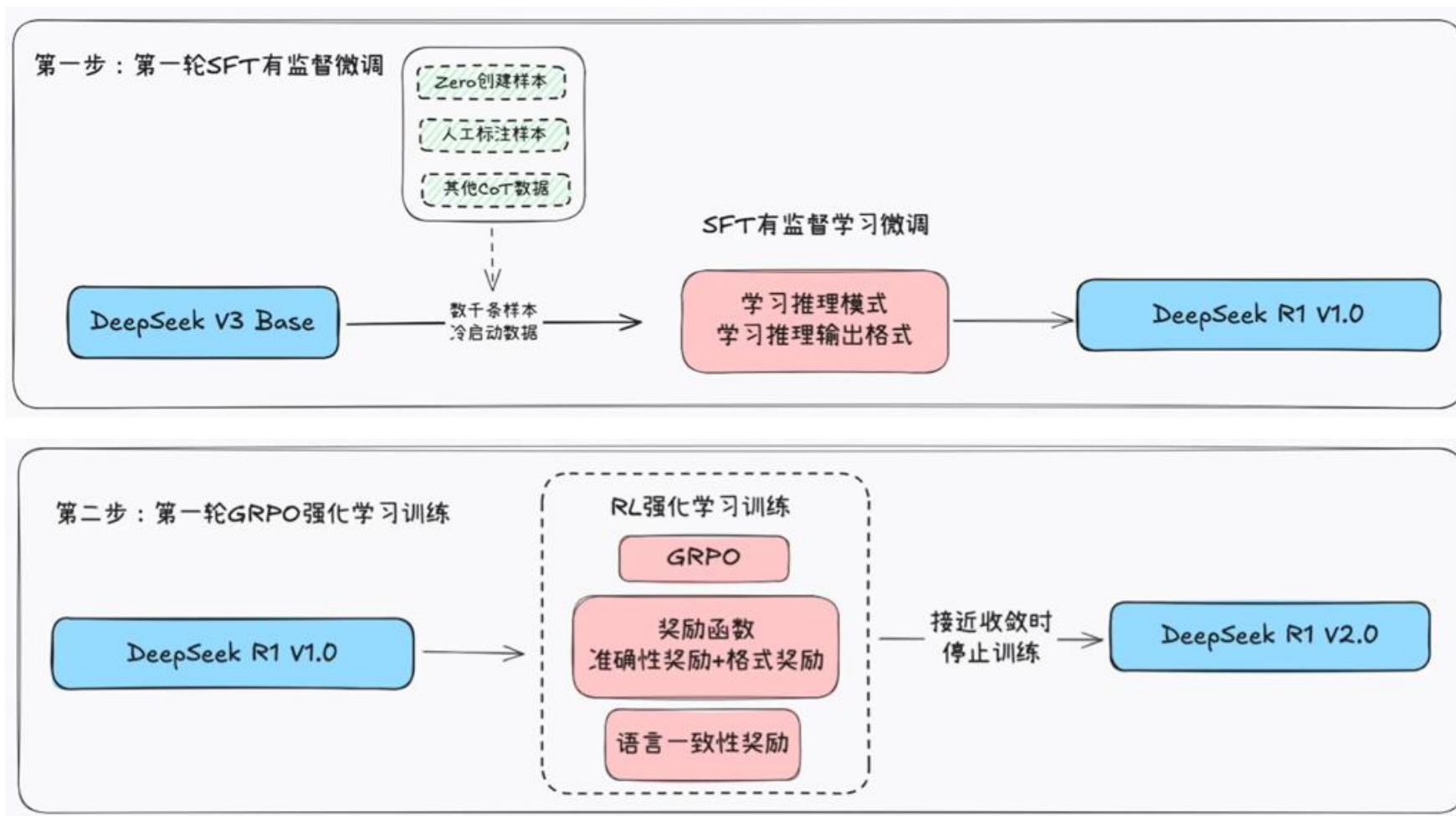
\dots

Table 3 | An interesting “aha moment” of an intermediate version of DeepSeek-R1-Zero. The model learns to rethink using an anthropomorphic tone. This is also an aha moment for us, allowing us to witness the power and beauty of reinforcement learning.

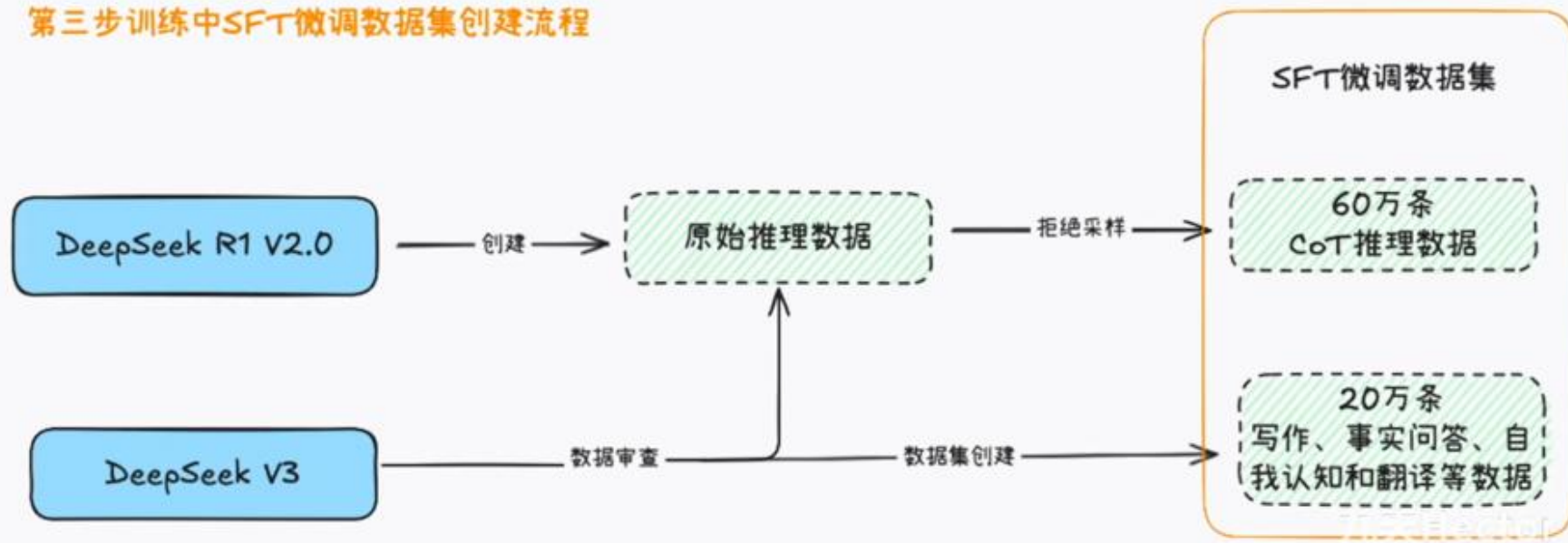
2. 训练DeepSeek R1 (4步)



2. 训练DeepSeek R1 (4步)



第三步训练中SFT微调数据集创建流程



第三步：第二轮SFT有监督微调



第四步：第二轮RL强化学习训练

DeepSeek R1 V3.0

RL强化学习训练

GRPO

推理数据 (Reasoning Data)
通用数据 (General Data)

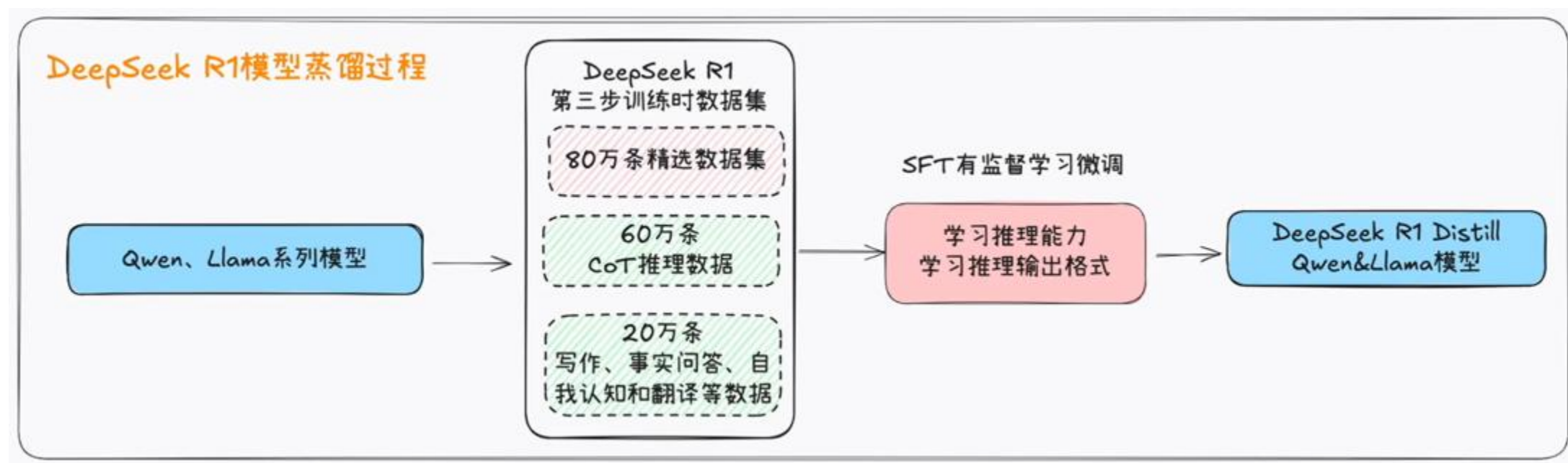
有用性 (Helpfulness)
无害性 (Harmlessness)

训练得到

DeepSeek R1

去中心化

3. 模型蒸馏



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.

效果很好，蒸馏后较小规模的模型能超过原始更大规模的模型，且蒸馏的32B和70B在大多数测试都能超过o1-mini。

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	90.8
	MMLU-Redux (EM)	88.9	88.0	89.1	86.7	-	92.9
	MMLU-Pro (EM)	78.0	72.6	75.9	80.3	-	84.0
	DROP (3-shot F1)	88.3	83.7	91.6	83.9	90.2	92.2
	IF-Eval (Prompt Strict)	86.5	84.3	86.1	84.8	-	83.3
	GPQA Diamond (Pass@1)	65.0	49.9	59.1	60.0	75.7	71.5
	SimpleQA (Correct)	28.4	38.2	24.9	7.0	47.0	30.1
	FRAMES (Acc.)	72.5	80.5	73.3	76.9	-	82.5
	AlpacaEval2.0 (LC-winrate)	52.0	51.1	70.0	57.8	-	87.6
	ArenaHard (GPT-4-1106)	85.2	80.4	85.5	92.0	-	92.3
Code	LiveCodeBench (Pass@1-COT)	38.9	32.9	36.2	53.8	63.4	65.9
	Codeforces (Percentile)	20.3	23.6	58.7	93.4	96.6	96.3
	Codeforces (Rating)	717	759	1134	1820	2061	2029
	SWE Verified (Resolved)	50.8	38.8	42.0	41.6	48.9	49.2
	Aider-Polyglot (Acc.)	45.3	16.0	49.6	32.9	61.7	53.3
Math	AIME 2024 (Pass@1)	16.0	9.3	39.2	63.6	79.2	79.8
	MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4	97.3
	CNMO 2024 (Pass@1)	13.1	10.8	43.2	67.6	-	78.8
Chinese	CLUEWSC (EM)	85.4	87.9	90.9	89.9	-	92.8
	C-Eval (EM)	76.7	76.0	86.5	68.9	-	91.8
	C-SimpleQA (Correct)	55.4	58.7	68.0	40.3	-	63.7

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

- 未来计划要提升DeepSeek-R1的下面几个能力
- 通用能力：目前，DeepSeek-R1在函数调用、多轮对话、复杂角色扮演和JSON输出等任务上的能力不如DeepSeek-V3。未来将探索CoT能够在这些任务上提升多少。
- 语言混合：DeepSeek-R1目前针对中文和英文进行了优化，这可能导致处理其他语言查询时出现语言混合问题。例如，即使查询是非英语或非中文，DeepSeek-R1也可能使用英文进行推理和回应。我们计划在未来的更新中解决这一限制。
- 提示工程：在评估DeepSeek-R1时，我们观察到它对提示词敏感。few shot会降低其性能。因此，我们建议用户直接描述问题，并使用zero shot指定输出格式以获得最佳结果。
- 软件工程任务（类似cursor的代码编辑、代码审查等方面）：由于评估时间较长，影响了强化学习过程的效率，大规模强化学习尚未广泛应用于软件工程任务。因此，DeepSeek-R1在软件工程基准测试上并未显示出相较于DeepSeek-V3的巨大改进。未来版本将通过在软件工程数据上实施拒绝采样或在强化学习过程中结合异步评估来提高效率，从而解决此问题。