





















A deep seek of DeepSeek Models

李坤泽

目录

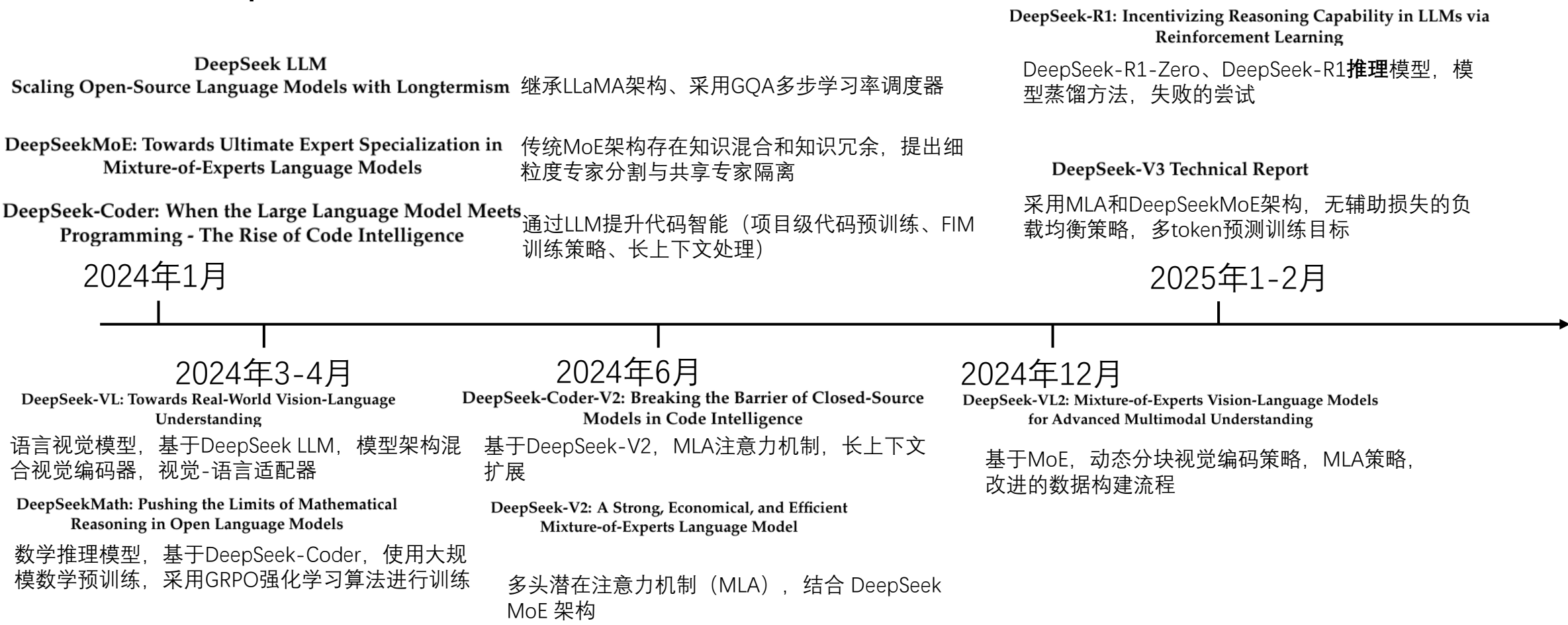
- DeepSeek系列论文回顾
- DeepSeekR1细节深入
- Deepseek本地部署

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

DeepSeek系列论文

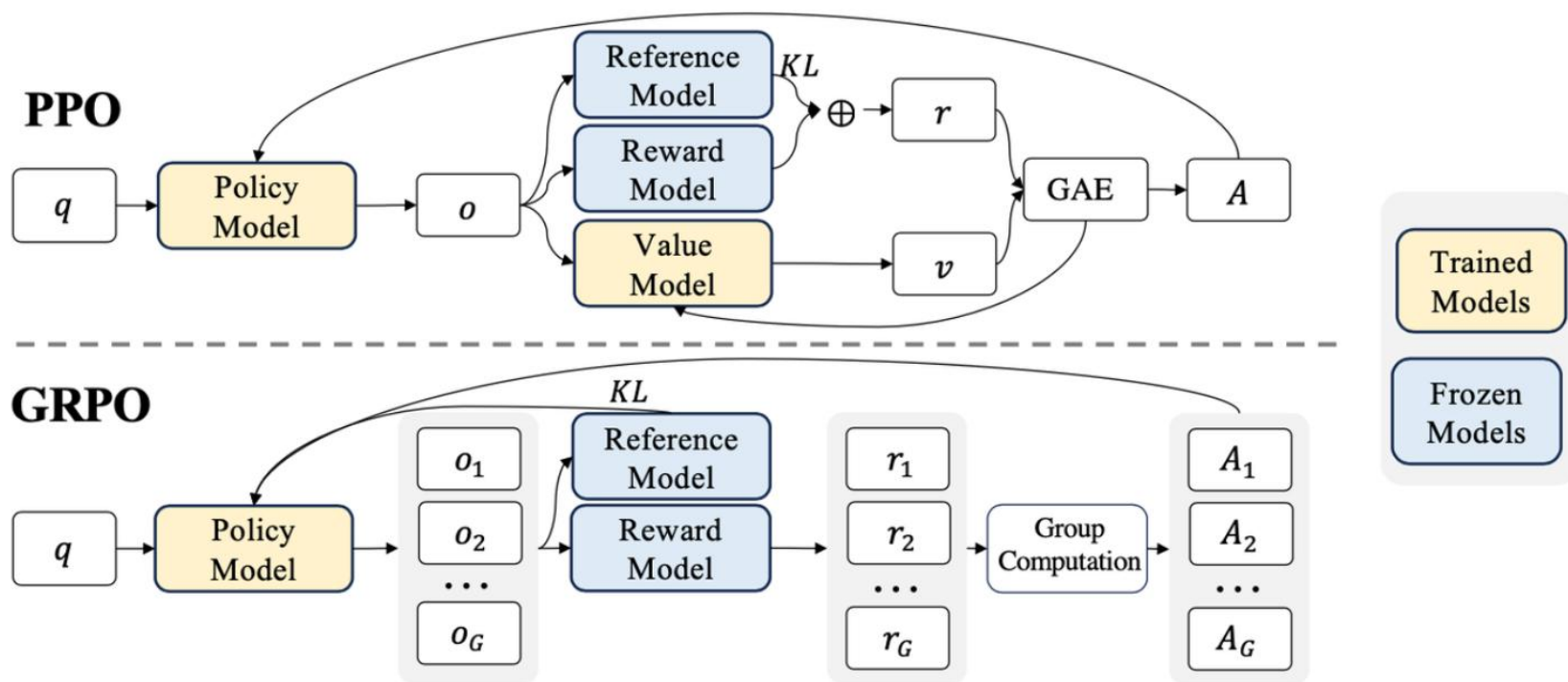


总览

- V1 (GQA) → V2 (混合专家架构、 MLA) → V3 (负载均衡、 **MTP**) → R1 (强化学习驱动)
- Coder (2024.1) → Coder V2 (2024.6)
- VL (2024.3) → Math (2024.4) → VL-2 (2024.8)
- 效率优先策略 (MoE、 RL)
- 开源生态布局 (语言模型、代码模型)
- 跨领域技术融合 (视觉-语言, mathRL)

GRPO Group Relative Policy Optimization

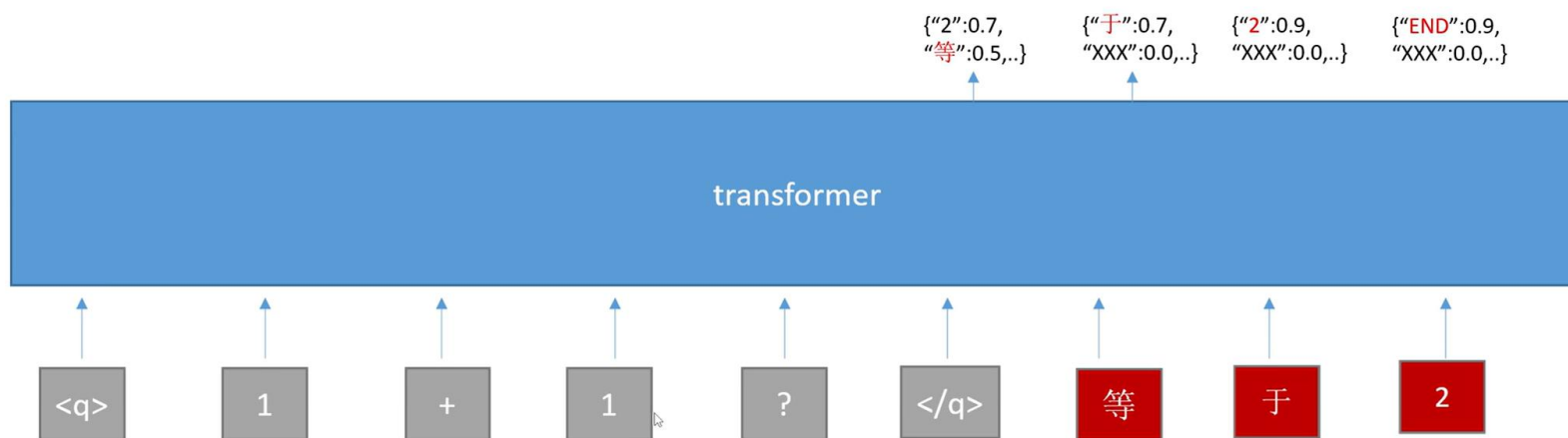
- 使用同一问题下多个采样输出的平均奖励作为基线



GRPO-采样

Next token probs的top K随机选，形成多样性

END



第一组

Query: <q>1+1?</q>

Completion: 2

Full: <q>1+1?2</q>2

第二组

Query: <q>1+1?</q>

Completion: 等于2

Full: <q>1+1?</q>等于2

GRPO-reward

自行设计，表达你认可的回答形式

第一组

Query: `<q>1+1?</q>`

Completion: 2

Full: `<q>1+1?2</q>2`



0.5分

第二组

Query: `<q>1+1?</q>`

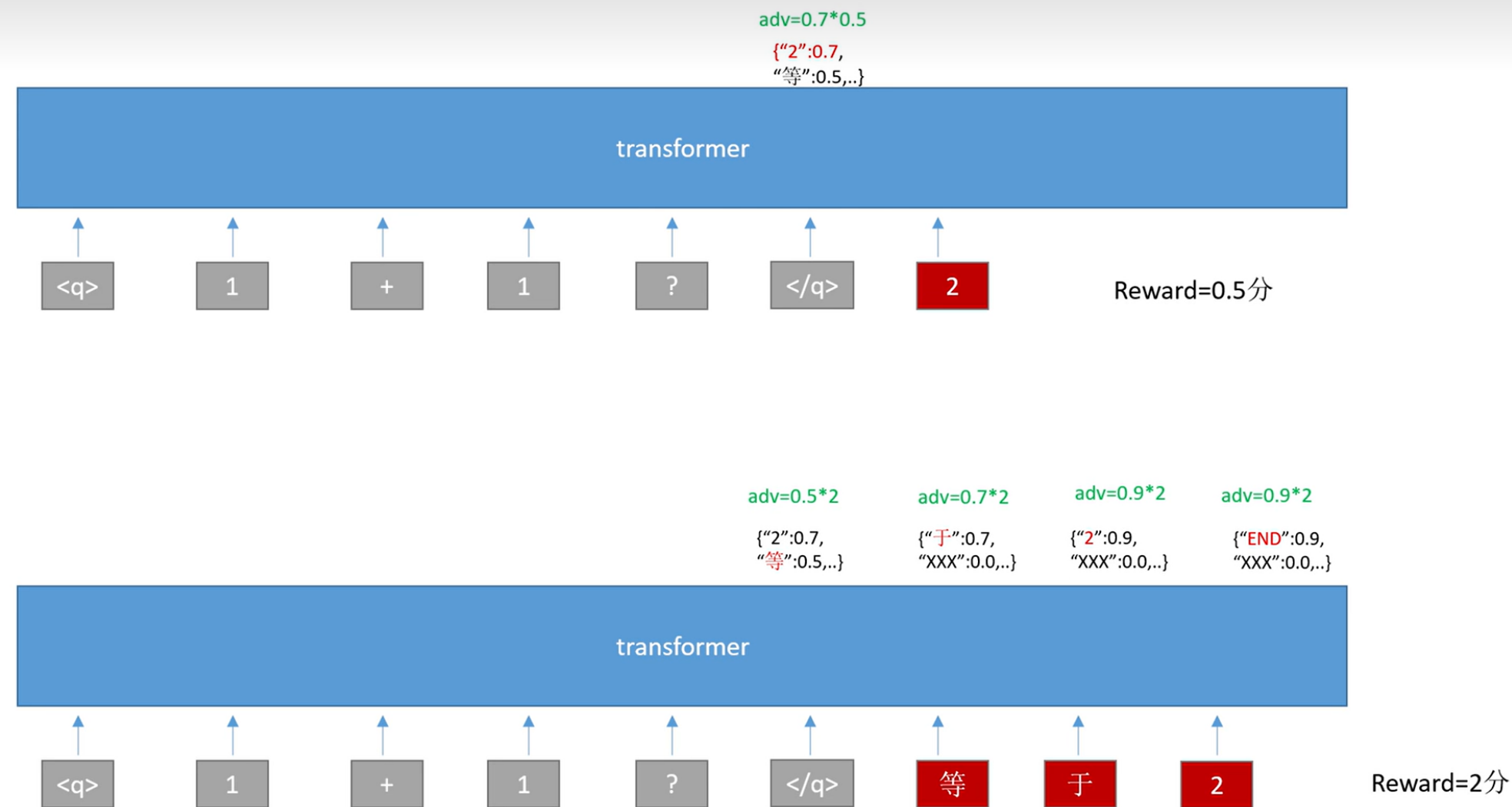
Completion: 等于2

Full: `<q>1+1?</q>等于2`



2分

GRPO-advantage



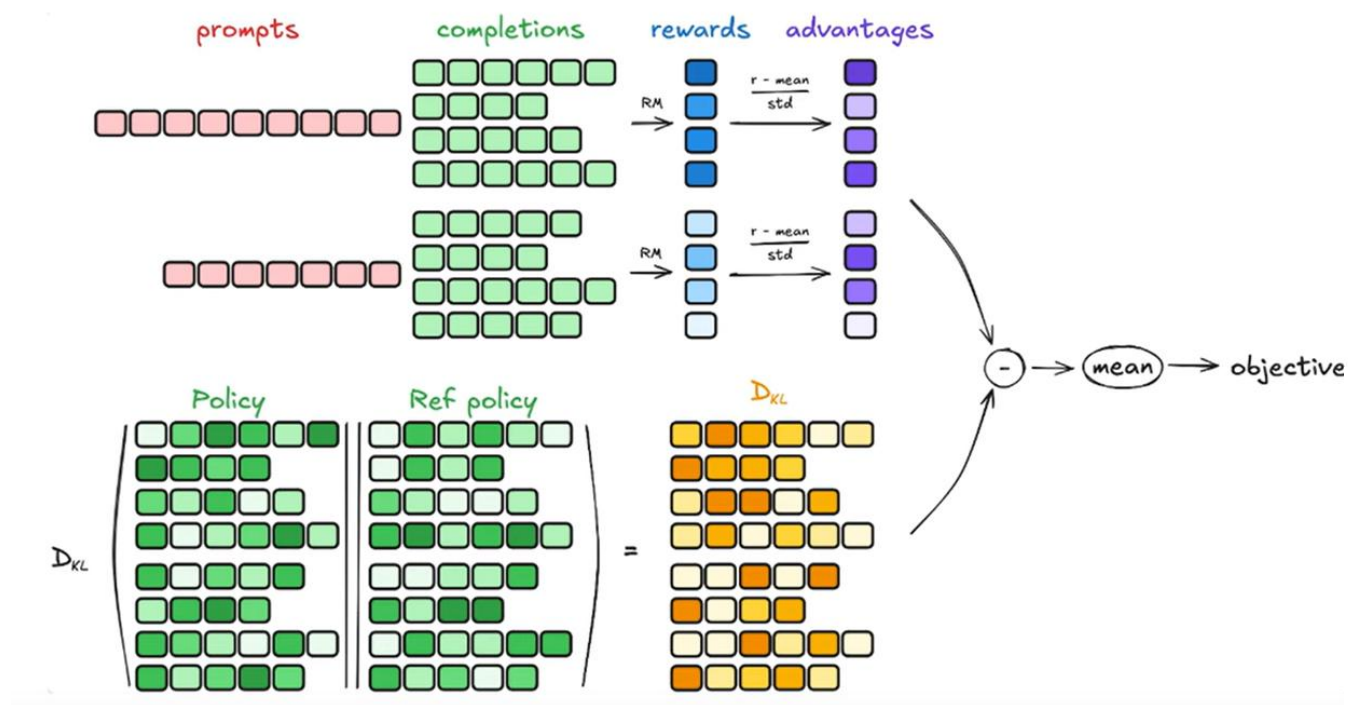
GRPO-loss

$$\text{Loss} = - \text{Sum}(\text{adv}=0.7*0.5 \quad \text{adv}=0.5*2 \quad \text{adv}=0.7*2 \quad \text{adv}=0.9*2 \quad \text{adv}=0.9*2) / 5$$



《原理》
希望reward高的
Next token，其
prob能继续走高。
这样就符合人类
的偏好了。

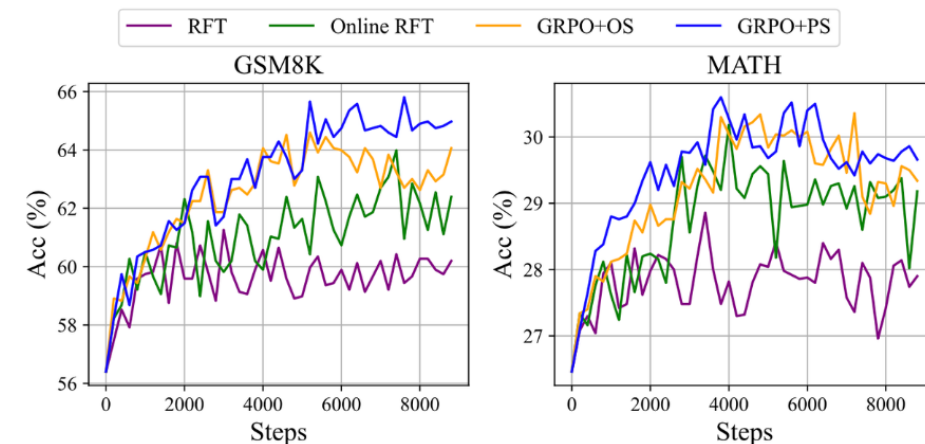
GRPO-loss-防止RL过度优化



$$\hat{A}_{i,t} = \frac{r_i - \text{mean}(\mathbf{r})}{\text{std}(\mathbf{r})}$$

$$\mathbb{D}_{KL}[\pi_{\theta} \parallel \pi_{\text{ref}}] = \frac{\pi_{\text{ref}}(o_{i,t} \mid q, o_{i,<t})}{\pi_{\theta}(o_{i,t} \mid q, o_{i,<t})} - \log \frac{\pi_{\text{ref}}(o_{i,t} \mid q, o_{i,<t})}{\pi_{\theta}(o_{i,t} \mid q, o_{i,<t})} - 1,$$

$$\mathcal{L}_{\text{GRPO}}(\theta) = -\frac{1}{G} \sum_{i=1}^G \frac{1}{|o_i|} \sum_{t=1}^{|o_i|} \left[\frac{\pi_{\theta}(o_{i,t} \mid q, o_{i,<t})}{[\pi_{\theta}(o_{i,t} \mid q, o_{i,<t})]_{\text{no grad}}} \hat{A}_{i,t} - \beta \mathbb{D}_{KL}[\pi_{\theta} \parallel \pi_{\text{ref}}] \right]$$

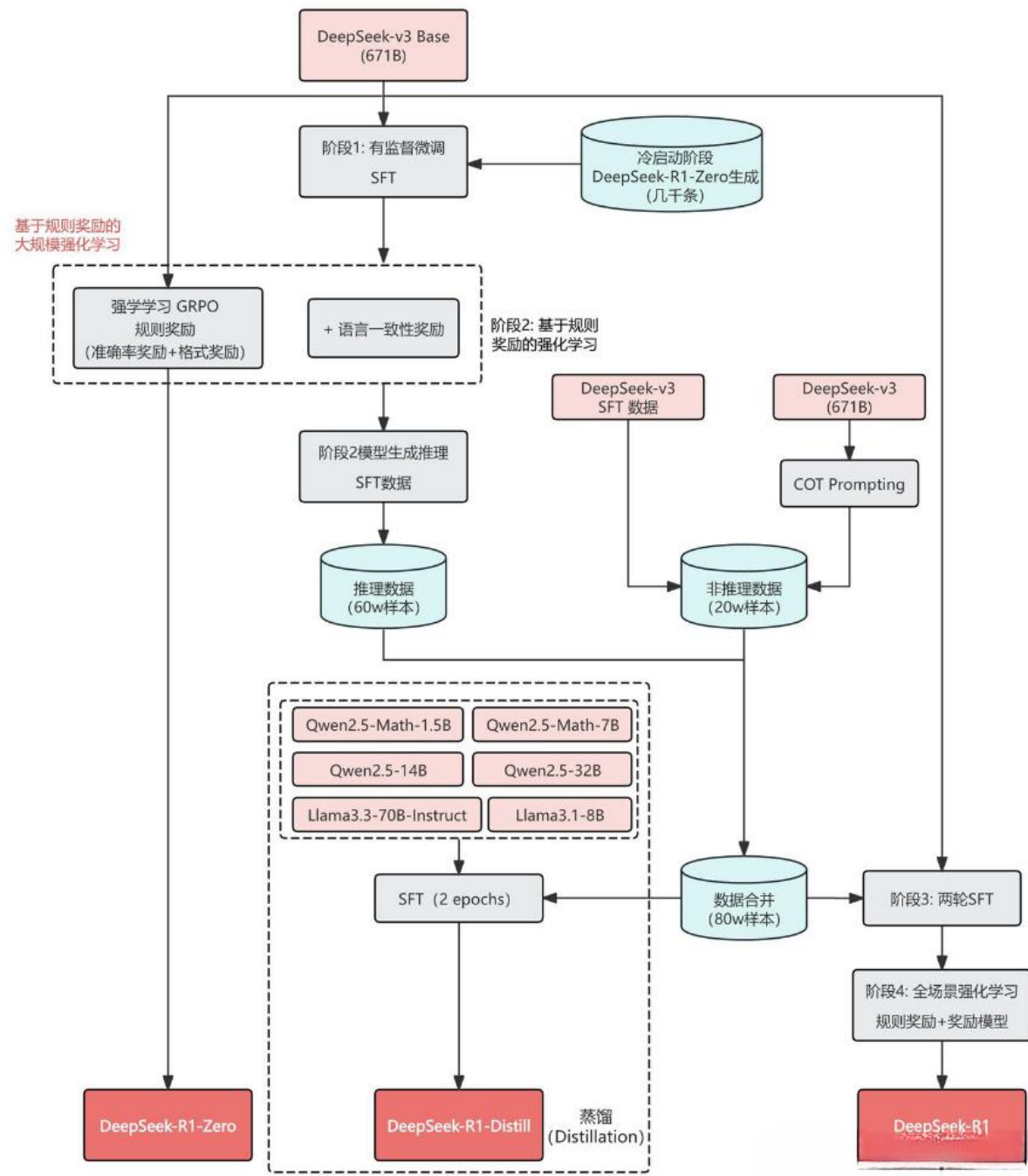


V3 vs R1

- DeepSeek V3
 - 为通用自然语言处理模型，采用混合专家（MoE）架构，参数总量达 6710 亿。其优势在于高效处理多模态任务（文本、图像、音频等）和长文本处理能力（支持 128K 上下文窗口），适用于内容生成、多语言翻译、智能客服等场景。
 - MoE
 - MLA
 - 负载均衡
- DeepSeek R1
 - 专注于复杂逻辑推理任务，基于强化学习（RL）训练，无需大量监督微调（SFT）。擅长数学证明、代码生成、决策优化等场景。其独特之处在于输出答案前展示“思维链”，增强透明度和可信度。
 - 冷启动
 - 强化学习

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

- 开源 DeepSeek-R1-Zero, 预训练模型直接 RL, 不走 SFT。
- 开源 DeepSeek-R1 推理大模型, 与 o1 性能相近。
- 开源用 R1 数据蒸馏的 Qwen、Llama 系列小模型, 蒸馏模型超过 o1-mini



1. 训练DeepSeek R1 Zero

第一阶段：训练DeepSeek R1 Zero



R1-zero性能

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.

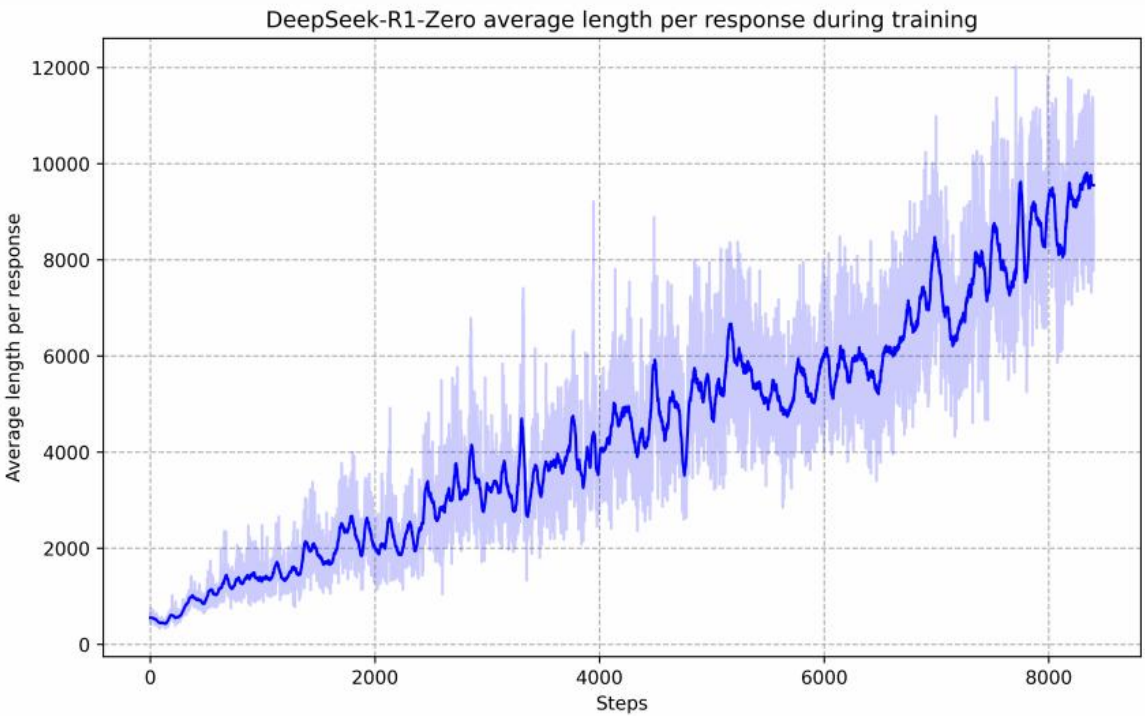
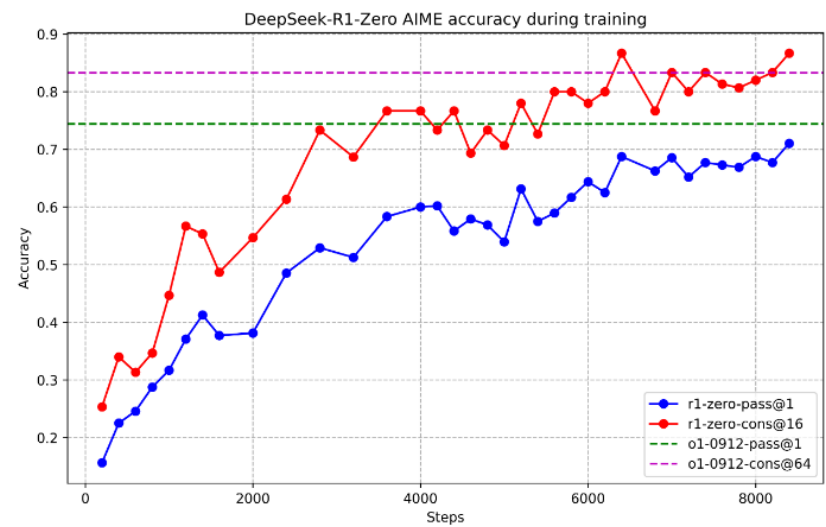


Figure 2 | AIME accuracy of DeepSeek-R1-Zero during training. For each question, we sample 16 responses and calculate the overall average accuracy to ensure a stable evaluation.

R1-zero-aha moment

-问题：如果 $a > 1$ ，那么 $\sqrt{a} - \sqrt{a+x} = x$ 的实数解之和等于 -

答：<think>为了解方程 $\sqrt{a} - \sqrt{a+x} = x$ ，我们首先两边平方...

$$(\sqrt{a} - \sqrt{a+x})^2 = x^2 \Rightarrow a - \sqrt{a+x} = x^2.$$

重新排列以隔离内部平方根项：

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

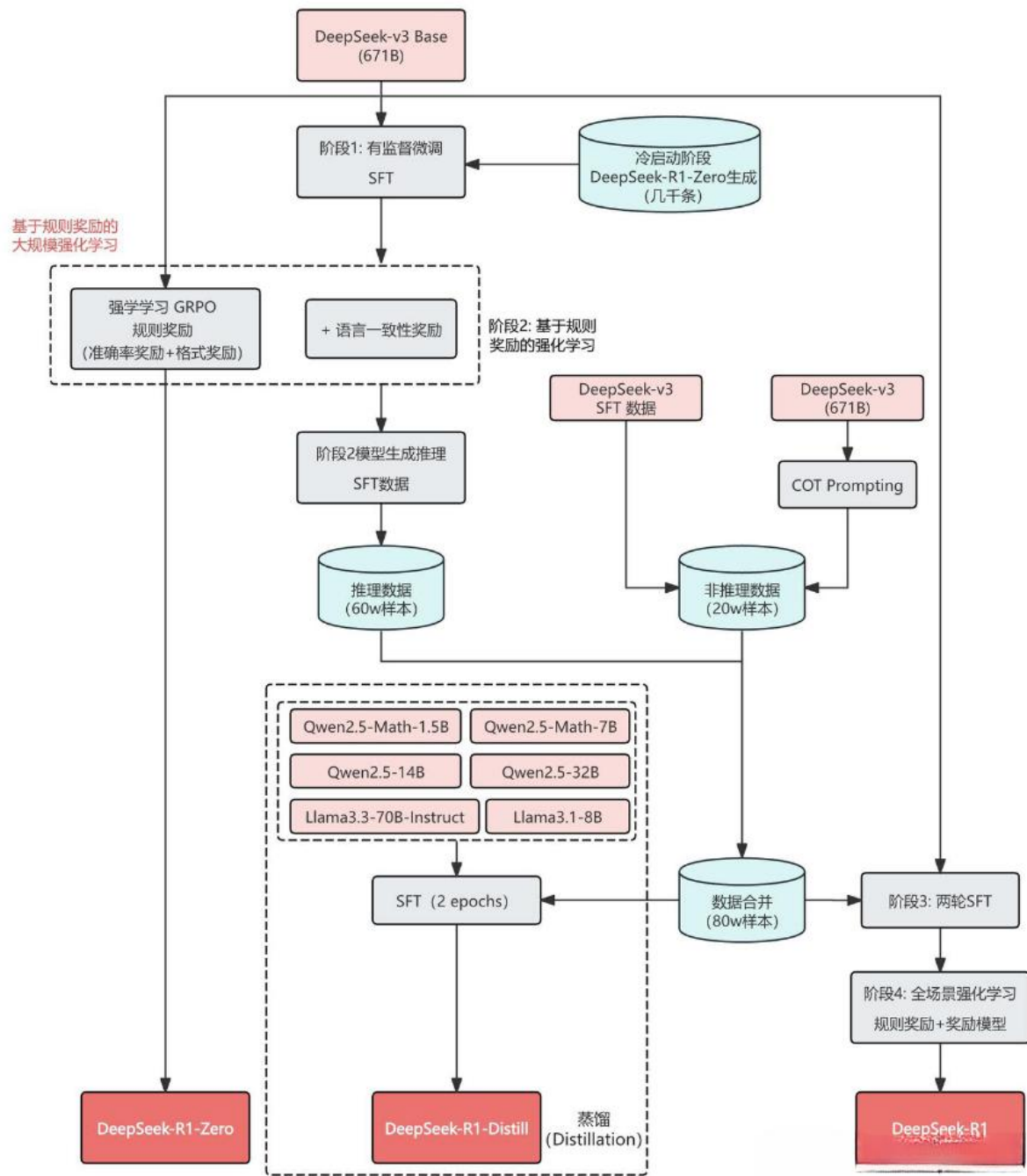
等一下，等一下。等一下。这里有一个我需要注意的地方。让我们一步一步

重新评估这一步，以确定正确的和是否可以... 我们首先从方程开始：

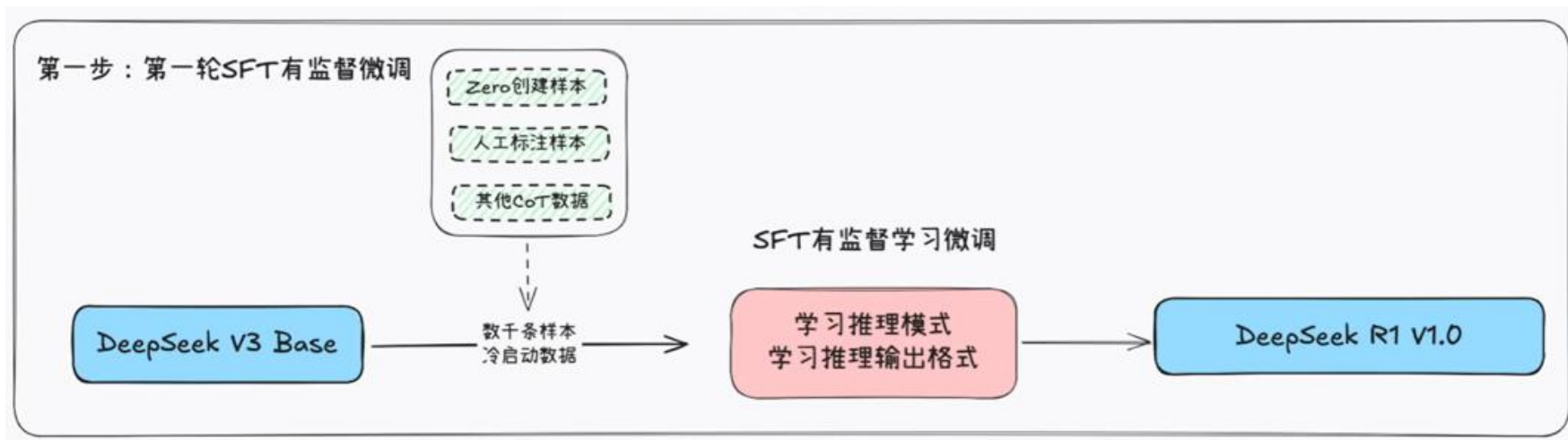
$\sqrt{a} - \sqrt{a+x} = x$ 首先，让我们平方两边：

$a - \sqrt{a+x} = x^2 \Rightarrow \sqrt{a+x} = a - x^2$ 接下来，我可以再次平方两边，处理方程：...

2. 训练DeepSeek R

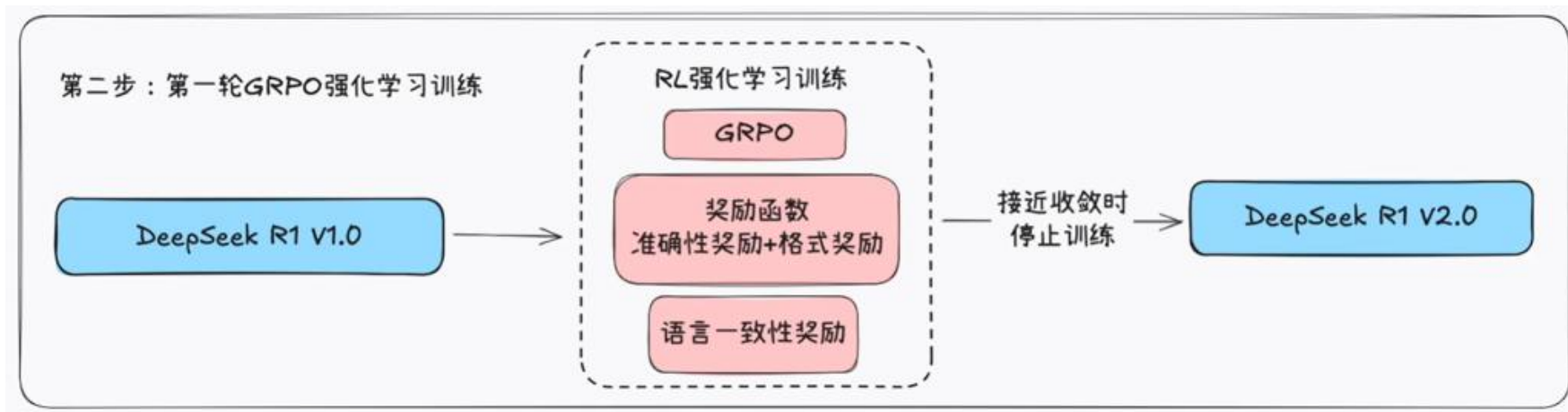


2. 训练DeepSeek R1（第1步）

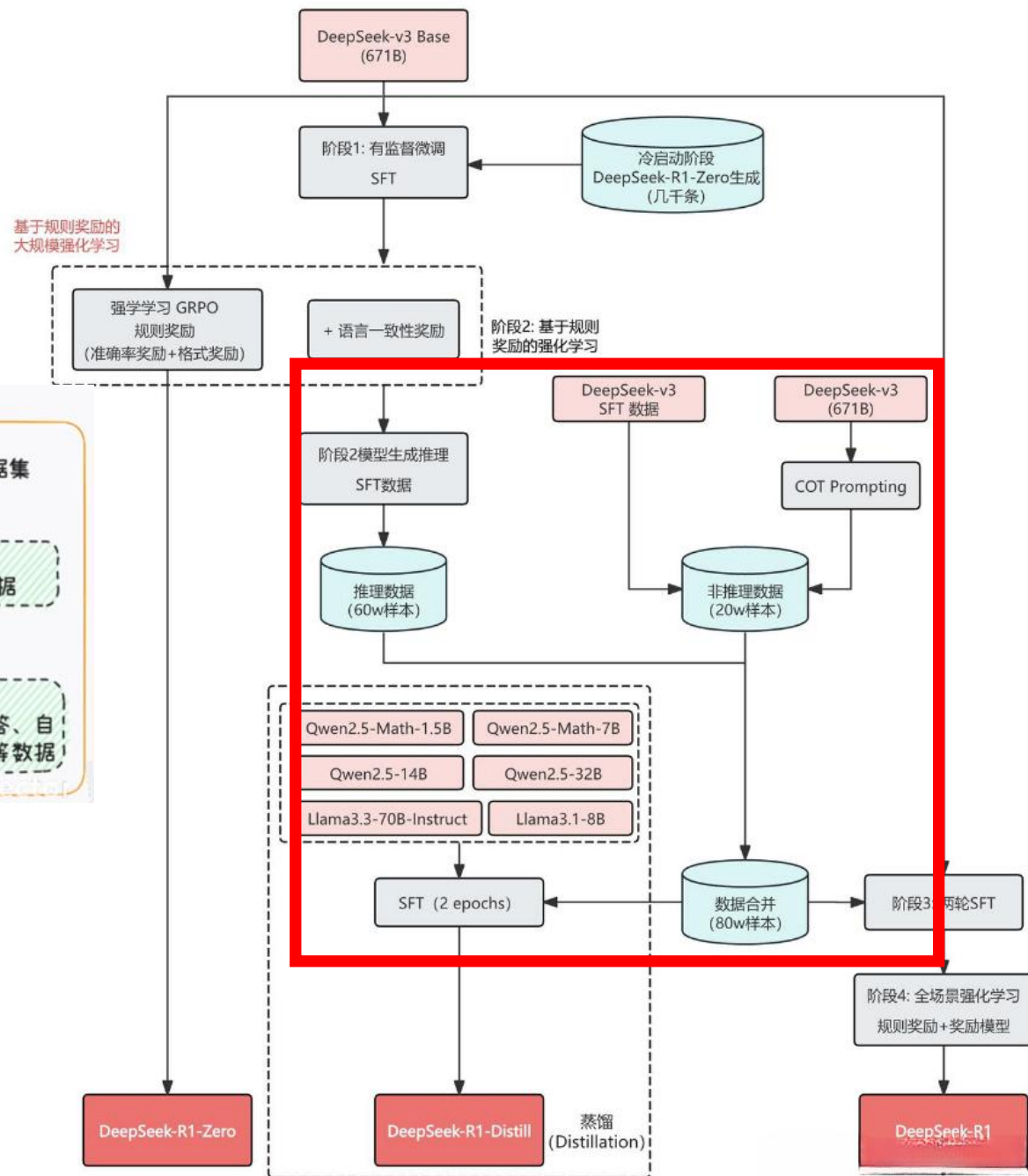


|special_token| <reasoning_process> |special_token| <summary>

2. 训练DeepSeek R1（第2步）



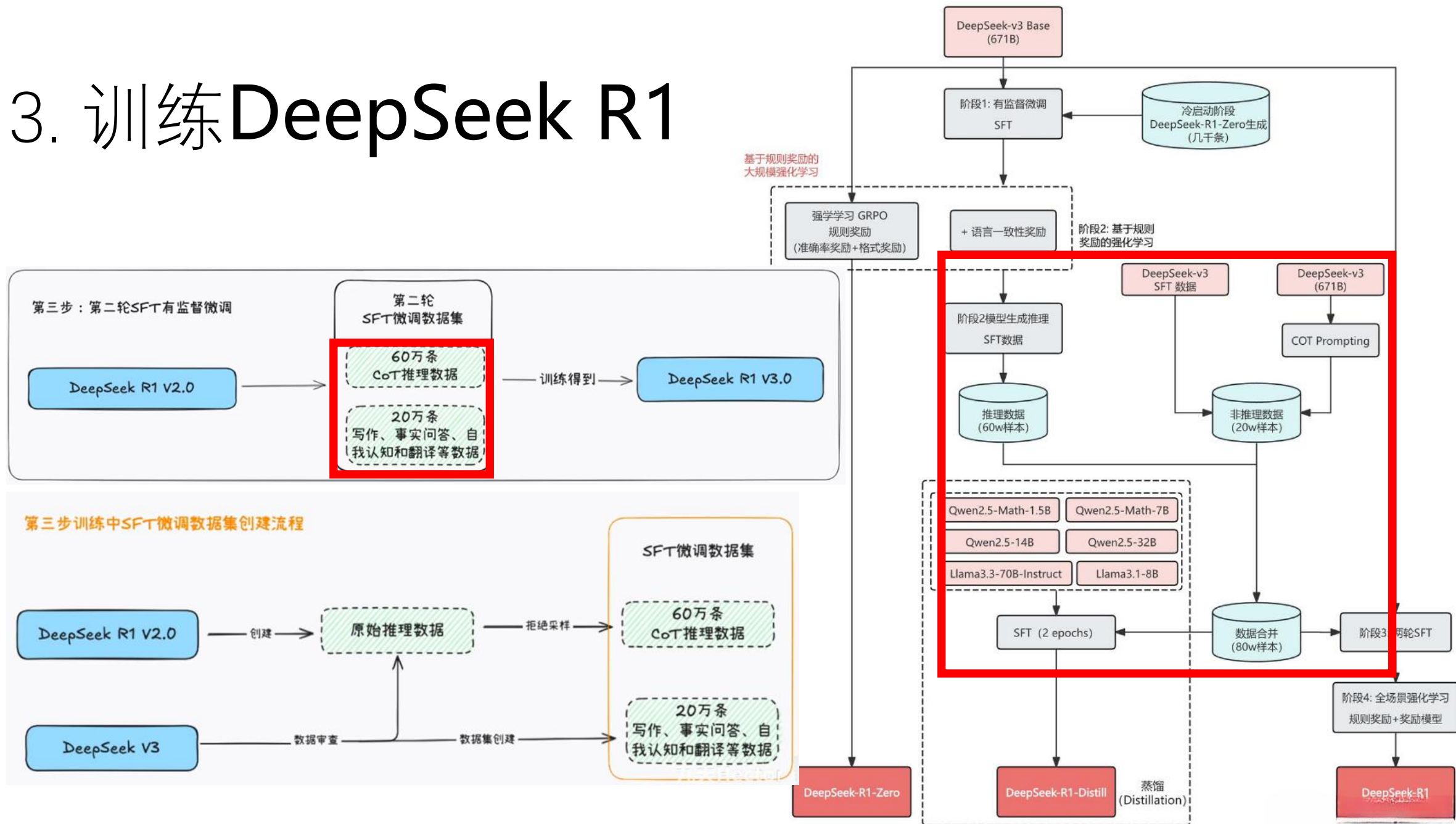
第三步训练中SFT微调数据集创建流程



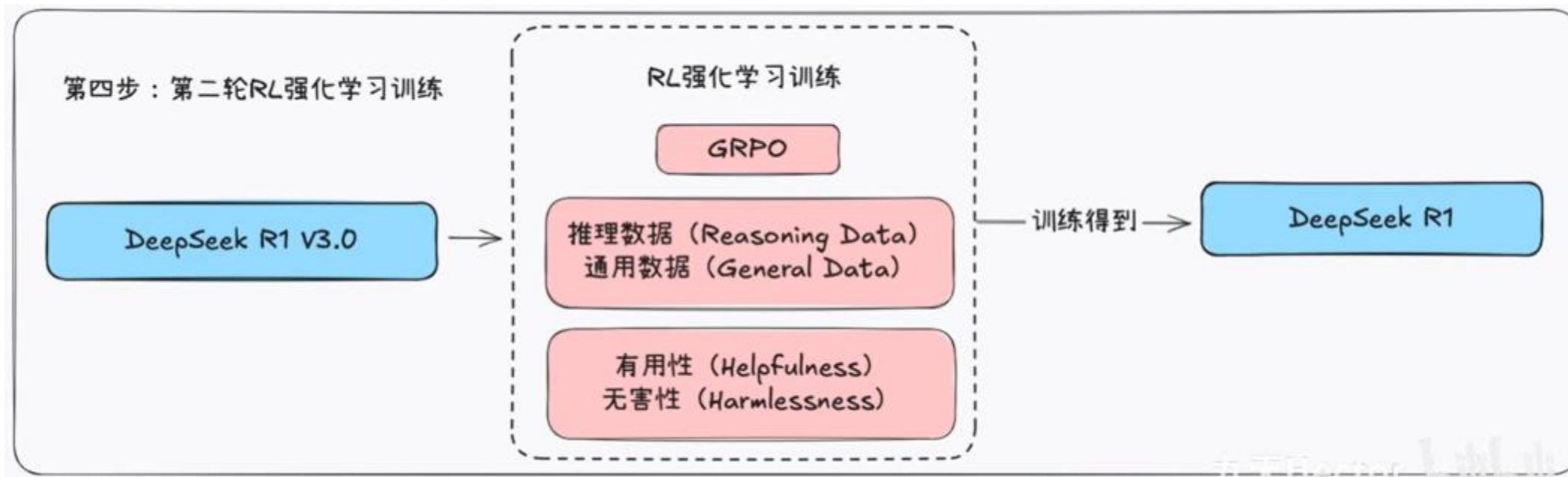
拒绝采样

- 步骤 1: 选择提议分布 $q(x)$
 - 提议分布 $q(x)$ 满足以下条件:
 - 易于采样。
 - 存在一个常数 M , 使得 $M \cdot q(x) \geq p(x)$ 对所有 x 成立。
- 步骤 2: 生成候选样本
 - 从提议分布 $q(x)$ 中生成一个候选样本 x 。
- 步骤 3: 计算接受概率
 - 计算接受概率 $A(x)$:
 - $A(x) = p(x) / M \cdot q(x)$, $p(x)$ 是目标分布, M 是满足 $M \cdot q(x) \geq p(x)$ 的常数。
- 步骤 4: 接受或拒绝样本
 - 生成一个均匀随机数 $u \sim \text{Uniform}(0,1)$ 。
 - 如果 $u \leq A(x)$, 则接受样本 x ; 否则, 拒绝样本并重新采样。

3. 训练DeepSeek R1



4. 训练DeepSeek R1（第4步）



基于规则的奖励

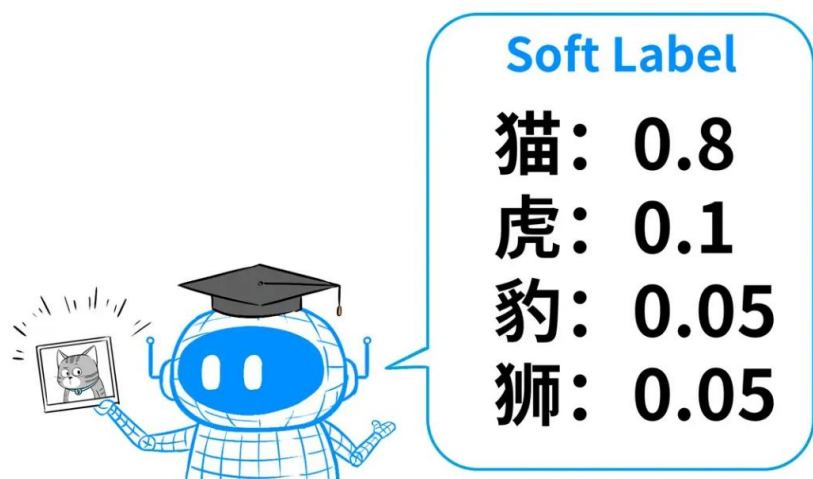
摘要

奖励模型

推理过程和摘要

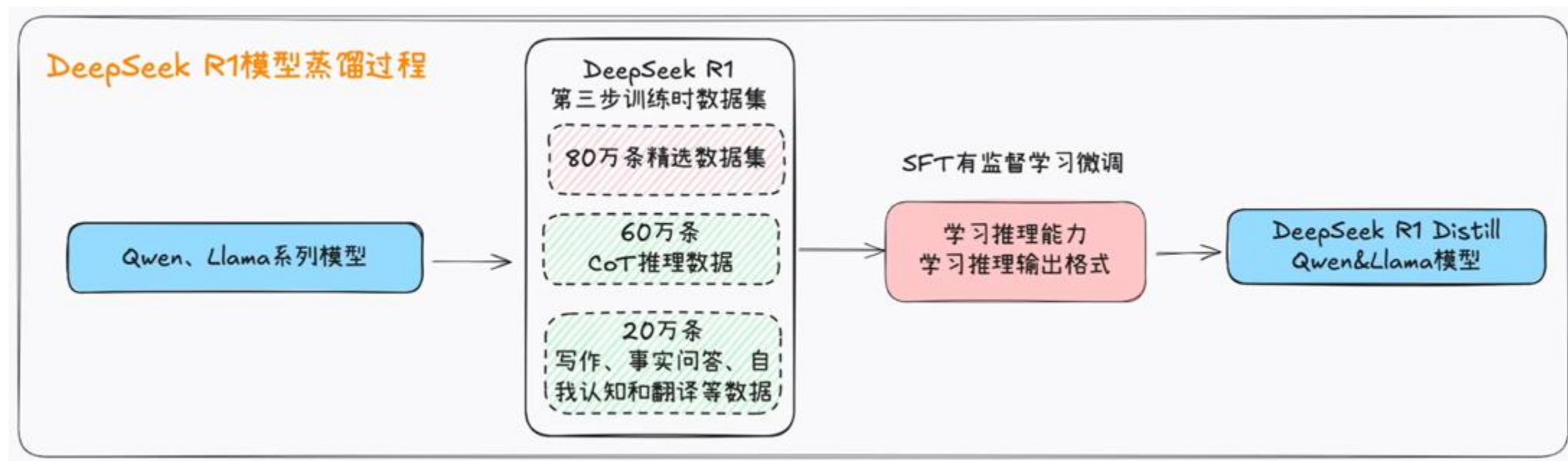
模型蒸馏

- 复杂模型（“教师模型”）的知识转移到简单模型（“学生模型”）
- 通过教师模型的输出（如概率分布）来指导学生模型的训练
- 训练教师模型——>生成软标签——>训练学生模型



$$\alpha \times (\text{蒸馏损失}) + (1-\alpha) \times (\text{真实监督损失})$$

模型蒸馏



失败的尝试

- 基于过程奖励模型 (PRM)
 - 推理步骤
 - 中间步骤的正确性
 - 奖励黑客
- 蒙特卡洛树搜索 (MCTS)
 - 搜索空间
 - 价值模型

实验设置

- 基准测试：数学推理（AIME 2024、MATH-500）、代码题（LiveCodeBench、Codeforces）、知识问答（MMLU、GPQA Diamond、SimpleQA）和开放生成场景（AlpacaEval2.0、ArenaHard）。对于蒸馏模型，AIME 2024、MATH-500、GPQA Diamond、Codeforces 和 LiveCodeBench
- 对比模型：评估了DeepSeek-V3、Claude-Sonnet-3.5-1022、GPT-4o-0513、OpenAI-o1-mini 和 OpenAI-o1-1217 等模型。

R1

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.

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。

蒸馏vsRL

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.

蒸馏 √ RL √

- 未来计划要提升DeepSeek-R1的下面几个能力
- **增强通用能力**：通过长链思维（CoT）提升模型在复杂任务（如函数调用、角色扮演）上的表现。
- **优化多语言支持**：解决语言混合问题，提升对非中英文语言的支持能力。
- **改进提示工程**：降低模型对提示的敏感性，提升在少样本提示下的表现。
- **提升软件工程任务性能**：通过拒绝采样和异步评估优化强化学习效率，增强模型在软件工程任务中的表现。

开源程度——开放权重

- - 完全开源：包含训练代码+数据+权重（如LLAMA2）
- - 开放权重：仅发布权重+推理代码（如早期的BERT）
- - 伪开源：仅提供API访问（GPT-3.5 和 GPT-4）

开源程度——开放权重

- <https://huggingface.co/deepseek-ai/DeepSeek-R1>
- <https://hf-mirror.com/deepseek-ai/DeepSeek-R1/tree/main>

The screenshot shows the Hugging Face Mirror interface. At the top, there's a search bar with the text "Search models, datasets, users...". Below the search bar, there are tabs for "Models", "Datasets", "Spaces", "Posts", "Docs", "Enterprise", and "Pricing". A language selector is set to "English". The main content area displays a list of models under the "deepseek" filter. The models listed include:

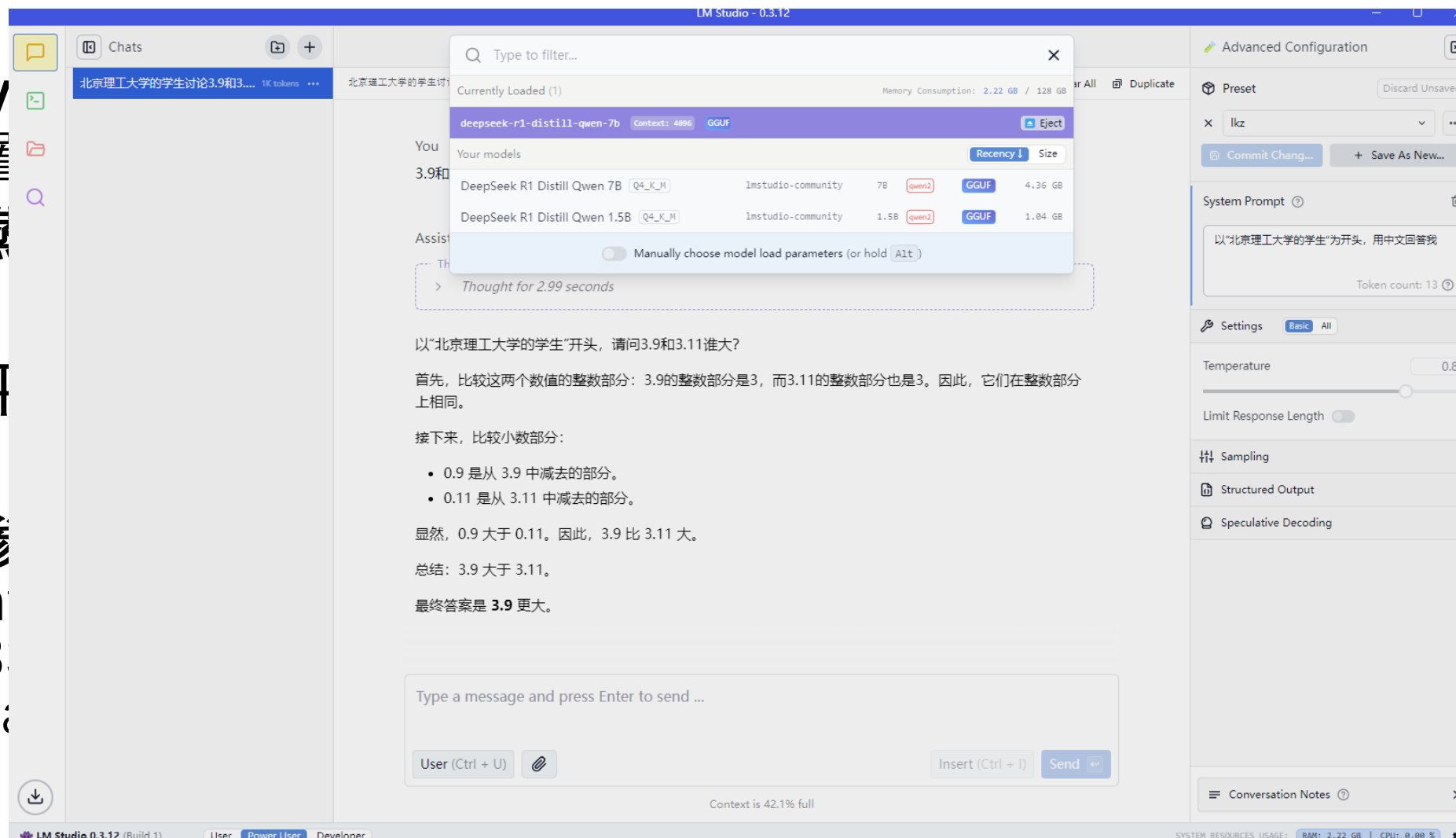
- deepseek-ai/DeepSeek-R1
- deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B
- deepseek-ai/Janus-Pro-7B
- deepseek-ai/DeepSeek-V3
- deepseek-ai/DeepSeek-R1-Distill-Qwen-32B
- nvidia/DeepSeek-R1-FP4
- huihui-ai/DeepSeek-671B-abliterated
- unsloth/DeepSeek-R1-GGUF
- deepseek-ai/DeepSeek-R1-Distill-Qwen-7B
- deepseek-ai/DeepSeek-R1-Distill-Llama-8B
- deepseek-ai/DeepSeek-R1-Distill-Llama-70B
- deepseek-ai/DeepSeek-R1-Distill-Qwen-14B
- meituan/DeepSeek-R1-Channel-INT8
- meituan/DeepSeek-R1-Block-INT8
- deepseek-ai/deepseek-v12-tiny
- mzadezmacher/DeepSeek-R1-Distill-Qwen-14B-Uncensored
- bartowski/DeepSeek-R1-Distill-Qwen-32B-GGUF
- cognitivecomputations/DeepSeek-R1-AWQ
- deepseek-ai/deepseek-v12
- deepseek-ai/DeepSeek-V3-Base

The screenshot shows the file structure of the DeepSeek-R1 repository. The repository is titled "deepseek-ai/DeepSeek-R1" and has 11k likes and 43.9k followers. The file structure is as follows:

- figures
- .gitattributes
- LICENSE
- README.md
- config.json
- configuration_deepseek.py
- generation_config.json
- model-0001-of-000163.safetensors
- model-0002-of-000163.safetensors
- model-0003-of-000163.safetensors
- model-0004-of-000163.safetensors
- model-0005-of-000163.safetensors
- model-0006-of-000163.safetensors
- model-0007-of-000163.safetensors
- model-0008-of-000163.safetensors
- model-0009-of-000163.safetensors
- model-0010-of-000163.safetensors

本地部署-交互式测试语言模型

- V
- 置
- 素
- 平
- 参
- h
- 3
- C

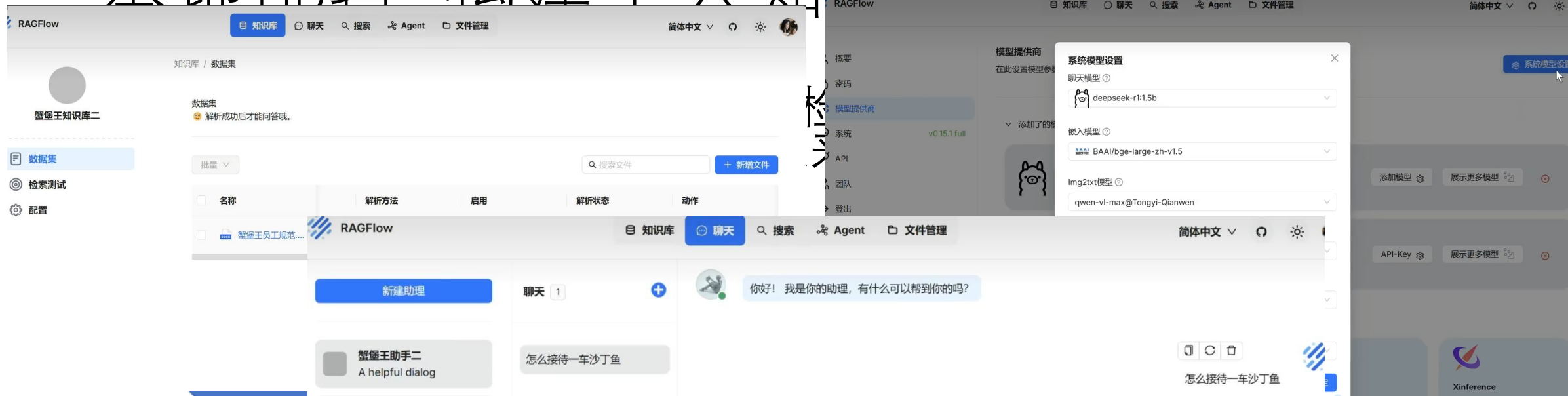


台配

=333.

39f

本地部署-构建个人知识库



• 流程: oll

• 参考:
<https://www.research-cai.com>

=333.337.s
!bf4f39f