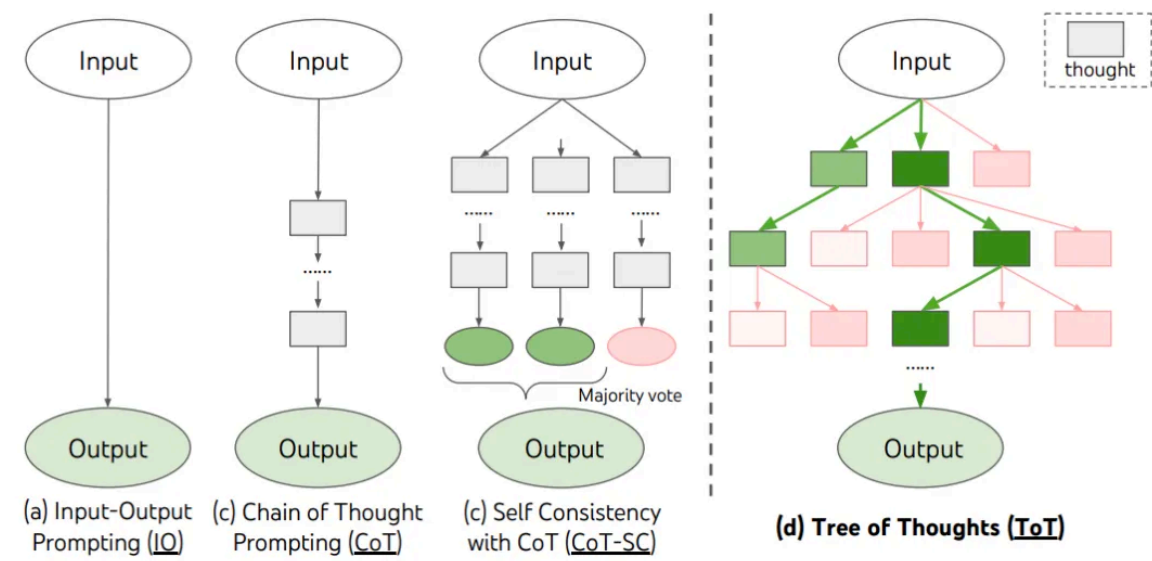


# DeepSeek-R1深度推理的核心技术：思维链（CoT）

原创 AllenTang 架构师带你玩转AI 2025年02月10日 22:11 湖北

DeepSeek的R1模型在处理逻辑推理、数学计算以及复杂问答等任务时，能够显式地展示其推理过程。R1深度推理的核心技术在于所采用的思维链（Chain of Thought, CoT）技术。思维链技术灵感来源于人类的思考方式，它要求模型将复杂问题逐步分解为多个简单步骤，并按照这些步骤逐一推导出最终答案。

通过这种分步推理的方法，R1模型的回答不仅更加精确可靠，而且其思考过程也变得清晰易懂。用户不再仅仅得到一个最终答案，而是能够跟随模型的思路，逐步理解它是如何得出这一结论的。



## 一、思维链（CoT）

**思维链（Chain of Thought, CoT）是什么？**思维链（CoT）是一种通过分步推理展示思考过程，增强大模型在复杂任务中推理能力和可解释性的技术。

思维链（CoT）的概念在论文《Chain-of-Thought Prompting Elicits Reasoning in Large Language Models》中被提出。

### Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wang Dale Schuurmans Maarten Bosma  
Brian Ichter Fei Xia Ed H. Chi Quoc V. Le Denny Zhou

Google Research, Brain Team  
{jasonwei,dennyzhou}@google.com

思维链的核心理念在于让模型像人类一样，将问题分解为多个步骤，逐步推导出答案。

1. 首先，思维链允许模型将多步骤问题分解为中间步骤，这意味着可以将额外的计算能力分配给需要更多推理步骤的问题。

- 其次，思维链为模型的行为提供了一个可解释的窗口，揭示了模型是如何得出特定答案的，并提供了在推理路径出错时进行调试的机会。
- 第三，思维链推理可用于数学文字题、常识推理和符号操作等任务，并且原则上至少可能适用于人类可以通过语言解决的任何任务。
- 最后，只需在少样本提示的示例中包含思维链序列的示例，就可以在足够大的现成语言模型中轻松引发思维链推理。

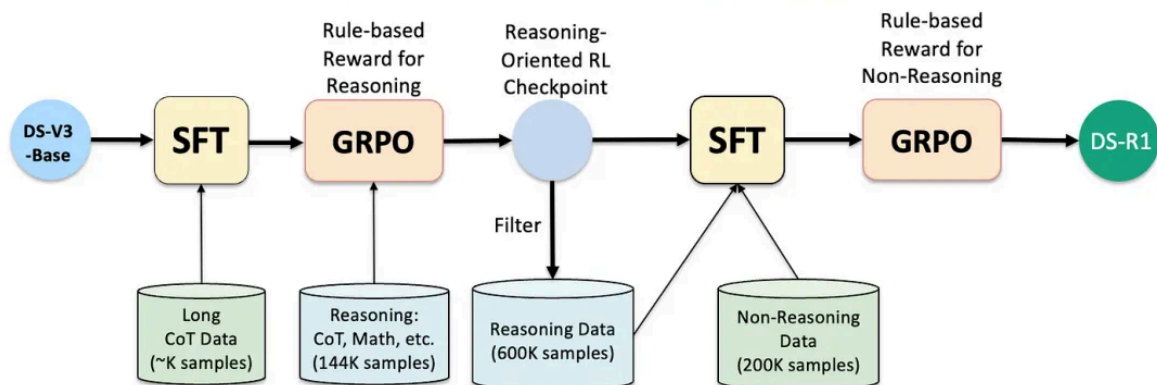
Chain-of-thought prompting has several attractive properties as an approach for facilitating reasoning in language models.

- First, chain of thought, in principle, allows models to decompose multi-step problems into intermediate steps, which means that additional computation can be allocated to problems that require more reasoning steps.
- Second, a chain of thought provides an interpretable window into the behavior of the model, suggesting how it might have arrived at a particular answer and providing opportunities to debug where the reasoning path went wrong (although fully characterizing a model's computations that support an answer remains an open question).
- Third, chain-of-thought reasoning can be used for tasks such as math word problems, commonsense reasoning, and symbolic manipulation, and is potentially applicable (at least in principle) to any task that humans can solve via language.
- Finally, chain-of-thought reasoning can be readily elicited in sufficiently large off-the-shelf language models simply by including examples of chain of thought sequences into the exemplars of few-shot prompting.

**DeepSeek-R1如何使用思维链（CoT）？** DeepSeek-R1首先让模型具有输出思维链的能力，然后再使用人类反馈强化学习（RLHF）和基于规则的推理奖励（Rule Base Reward for Reasoning）对模型进行强化学习的训练。

DeepSeek-R1模型在推理过程中就能够自然地生成思维链，并按照人类期望的方式进行推理。

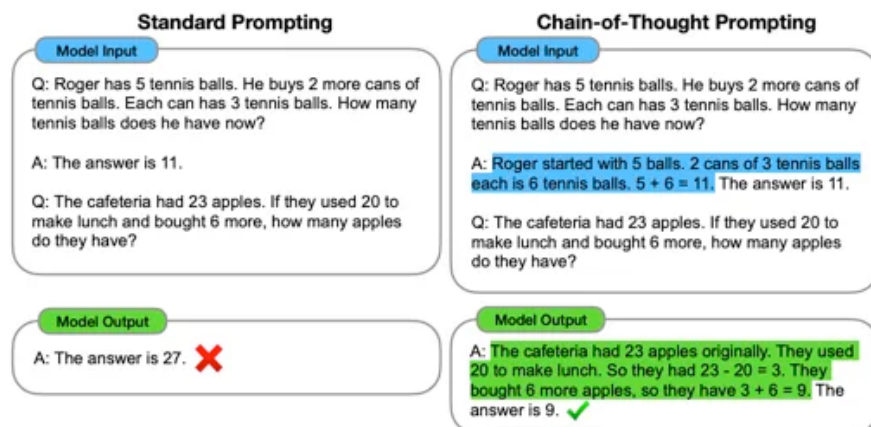
## DeepSeek-R1 Training Pipeline



### 二、Chain-of-Thought Prompting

**Chain-of-Thought Prompting是什么？** Chain-of-Thought Prompting是通过在提示中包含中间推理步骤的示例，引导大语言模型生成类似的思维链。

- 分步提示设计：**在编写大模型应用时，通过特定格式的提示（Prompt）要求模型展示思考过程。使用关键词如“Let's think step by step”触发逐步推理。支持零样本（Zero-shot）和少样本（Few-shot）两种模式。
- 中间状态生成：**模型生成中间推理步骤作为文本，每个步骤作为后续推理的上下文基础。通过自回归方式逐步生成内容。
- 结果提取：**从生成的文本中解析最终答案。通常以“Therefore, the answer is”等关键词标识结论。



如何让大语言模型（LLM）生成思维链？通过Chain-of-Thought Prompting，可以在少样本提示的示例中提供思维链推理的演示，让大语言模型能够生成思维链。

在少样本提示中，为每个示例添加与答案相关的思维链。这些示例应该清晰地展示如何从问题出发，通过一系列中间步骤，最终得出结论。

这样可以引导LLM在回答新问题时，也按照类似的思维链进行推理，通过这些包含中间推理步骤的示例，LLM能够学习到如何分解问题并生成思维链。

<p><b>Math Word Problems (free response)</b></p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. <math>5 + 6 = 11</math>. The answer is 11.</p>	<p><b>Math Word Problems (multiple choice)</b></p> <p>Q: How many keystrokes are needed to type the numbers from 1 to 500?</p> <p>Answer Choices: (a) 1156 (b) 1392 (c) 1480 (d) 1562 (e) 1788</p> <p>A: There are 9 one-digit numbers from 1 to 9. There are 90 two-digit numbers from 10 to 99. There are 401 three-digit numbers from 100 to 500. <math>9 + 90(2) + 401(3) = 1392</math>. The answer is (b).</p>	<p><b>CSQA (commonsense)</b></p> <p>Q: Sammy wanted to go to where the people were. Where might he go?</p> <p>Options: (a) race track (b) populated areas (c) desert (d) apartment (e) roadblock</p> <p>A: The answer must be a place with a lot of people. Race tracks, desert, apartments, and roadblocks don't have a lot of people, but populated areas do. So the answer is (b).</p>
<p><b>StrategyQA</b></p> <p>Q: Yes or no: Would a pear sink in water?</p> <p>A: The density of a pear is about <math>0.6 \text{ g/cm}^3</math>, which is less than water. Thus, a pear would float. So the answer is no.</p>	<p><b>Date Understanding</b></p> <p>Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?</p> <p>A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943.</p>	<p><b>Sports Understanding</b></p> <p>Q: Is the following sentence plausible? "Joao Moutinho caught the screen pass in the NFC championship."</p> <p>A: Joao Moutinho is a soccer player. The NFC championship is part of American football, not soccer. So the answer is no.</p>
<p><b>SayCan (instructing a robot)</b></p> <p>Human: How would you bring me something that isn't a fruit?</p> <p>Explanation: the user wants something to eat that isn't a fruit. An energy bar is not a fruit, so I will bring the user an energy bar.</p> <p>Plan: 1. find(energy bar) 2. pick(energy bar) 3. find(user) 4. put(energy bar) 5. done().</p>	<p><b>Last Letter Concatenation</b></p> <p>Q: Take the last letters of the words in "Lady Gaga" and concatenate them.</p> <p>A: The last letter of "Lady" is "y". The last letter of "Gaga" is "a". Concatenating them is "ya". So the answer is ya.</p>	<p><b>Coin Flip (state tracking)</b></p> <p>Q: A coin is heads up. Maybelle flips the coin. Shalonda does not flip the coin. Is the coin still heads up?</p> <p>A: The coin was flipped by Maybelle. So the coin was flipped 1 time, which is an odd number. The coin started heads up, so after an odd number of flips, it will be tails up. So the answer is no.</p>



DeepSeek 14

DeepSeek · 目录

上一篇

一文搞懂DeepSeek - DeepSeek的三种访问方式

下一篇

DeepSeek火爆背后的核心技术：模型压缩（知识蒸馏）