Hochschule Konstanz
Fakultät Informatik

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

Spotlight Talk, Seminar: Advanced Topics in Data Analysis and Deep Learning

Lukas Epple

About the paper

- Published 22 Jan 2025
- DeepSeek research Team (200 authors)

deepseek

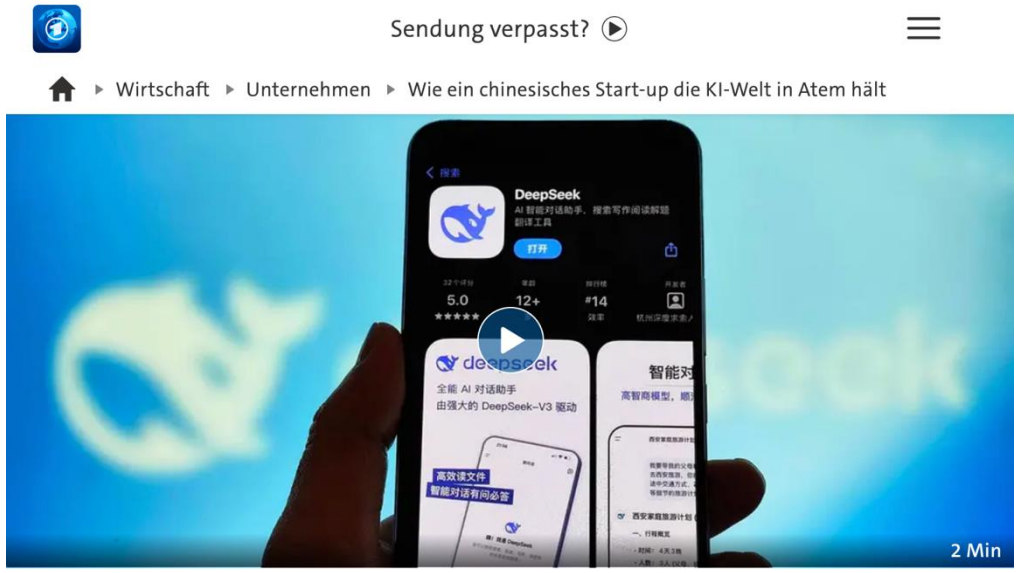
DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

DeepSeek-AI

research@deepseek.com

Abstract

We introduce our first-generation reasoning models, DeepSeek-R1-Zero and DeepSeek-R1. DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning (RL) without supervised fine-tuning (SFT) as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing reasoning behaviors. However, it encounters challenges such as poor readability, and language mixing. To address these issues and further enhance reasoning performance, we introduce DeepSeek-R1, which incorporates multi-stage training and cold-start data before RL. DeepSeek-R1 achieves performance comparable to OpenAI-o1-1217 on reasoning tasks. To support the research community, we open-source DeepSeek-R1-Zero, DeepSeek-R1, and six dense models (1.5B, 7B, 8B, 14B, 32B, 70B) distilled from DeepSeek-R1 based on Qwen and Llama.



HINTERGRUND Chinesisches KI-Start-up

DeepSeek, der Schrecken der US-Techgiganten

Stand: 28.01.2025 07:32 Uhr

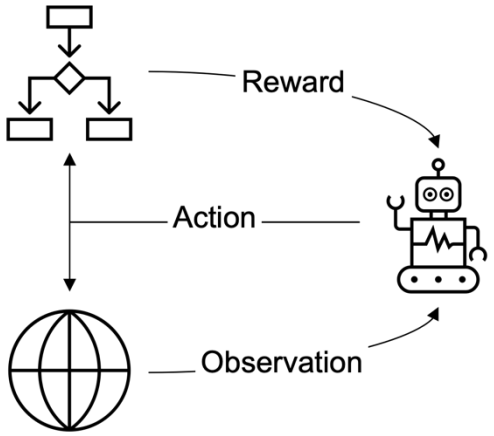
Das Start-up DeepSeek verblüfft und verunsichert die Tech-Welt. Denn das neueste KI-Modell des chinesischen Unternehmens soll deutlich effizienter sein als die Konkurrenz aus den USA.

https://www.tagesschau.de/wirtschaft/unternehmen/deepseek-ki-start-up-china-100.html

How can LLMs be encouraged to think reasonably in order to improve the quality of their outputs?

Context

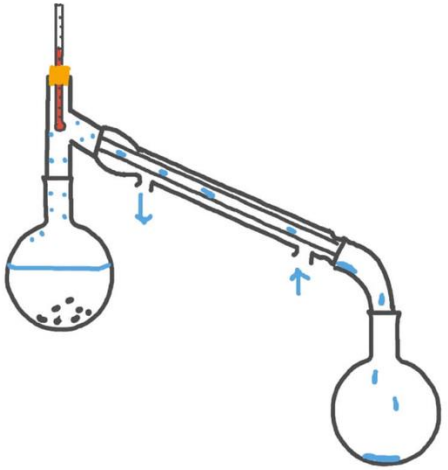
Improving LLMs



Reinforcement Learning

```
<think> reasoning process here </think>
<answer> final answer here </answer>
```

Reasoning („think slow“)



Knowledge Distillation

https://diwanegypt.com/wp-content/uploads/2020/08/9780141033570.jpg
https://media.nagwa.com/698132726417/en/thumbnail_l.jpeg

Whats special about the paper?

- First **open validation** of **pure* RL for reasoning**
  - No need for supervised training with lots of **labeled data** or other auxilliary tools
  - Opening doors for more **autonomous training** of models in the future

- Discovery of „**Aha moment**“
  - Not just imitating human CoT but developing **own thinking patterns**

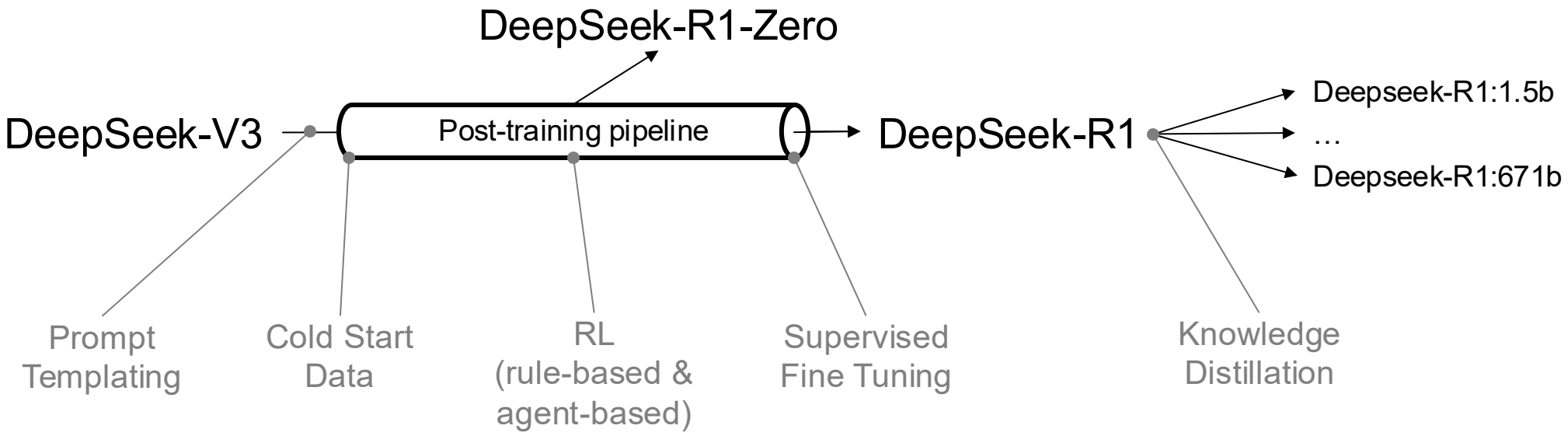
...
Wait, wait. Wait. That's an aha moment I can flag here.
Let's reevaluate this step-by-step to identify if the correct sum can be ...

- **Successful distillation** of reasoning capabilities

- Artifacts with **competitive performance**

https://i.ytimg.com/vi/rjevPB9epds/maxresdefault.jpg

Methodology



Thank you for your attention!

