The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity

Parshin Shojaee*† Iman Mirzadeh* Keivan Alizadeh
Maxwell Horton Samy Bengio Mehrdad Farajtabar

Apple

Abstract

Recent generations of frontier language models have introduced Large Reasoning Models (LRMs) that generate detailed thinking processes before providing answers. While these models demonstrate improved performance on reasoning benchmarks, their fundamental capabilities, scaling properties, and limitations remain insufficiently understood. Current evaluations primarily focus on established mathematical and coding benchmarks, emphasizing final answer accuracy. However, this evaluation paradigm often suffers from data contamination and does not provide insights into the reasoning traces' structure and quality. In this work, we systematically investigate these gaps with the help of controllable puzzle environments that allow precise manipulation of compositional complexity while maintaining consistent logical structures. This setup enables the analysis of not only final answers but also the internal reasoning traces, offering insights into how LRMs "think". Through extensive experimentation across diverse puzzles, we show that frontier LRMs face a complete accuracy collapse beyond certain complexities. Moreover, they exhibit a counter-intuitive scaling limit: their reasoning effort increases with problem complexity up to a point, then declines despite having an adequate token budget. By comparing LRMs with their standard LLM counterparts under equivalent inference compute, we identify three performance regimes: (1) low-complexity tasks where standard models surprisingly outperform LRMs, (2) medium-complexity tasks where additional thinking in LRMs demonstrates advantage, and (3) high-complexity tasks where both models experience complete collapse. We found that LRMs have limitations in exact computation: they fail to use explicit algorithms and reason inconsistently across puzzles. We also investigate the reasoning traces in more depth, studying the patterns of explored solutions and analyzing the models' computational behavior, shedding light on their strengths, limitations, and ultimately raising crucial questions about their true reasoning capabilities.

1 Introduction

Large Language Models (LLMs) have recently evolved to include specialized variants explicitly designed for reasoning tasks—Large Reasoning Models (LRMs) such as OpenAI's o1/o3 [1, 2], DeepSeek-R1 [3], Claude 3.7 Sonnet Thinking [4], and Gemini Thinking [5]. These models are new artifacts, characterized by their "*thinking*" mechanisms such as long Chain-of-Thought (CoT) with self-reflection, and have demonstrated promising results across various reasoning benchmarks. Their

*Equal contribution.
†Work done during an internship at Apple.
{p_shojaee, imirzadeh, kalizadehvahid, mchorton, bengio, farajtabar}@apple.com

《思考的錯覺》

APPLE 2025.6.7

「LLM看起來在思考，
其實只是在背答案。」

ProRL: Prolonged Reinforcement Learning Expands Reasoning Boundaries in Large Language Models

Mingjie Liu **Shizhe Diao** **Ximing Lu** **Jian Hu** **Xin Dong**
Yejin Choi **Jan Kautz** **Yi Dong**
NVIDIA
{mingjiel, sdiao, ximingl, jianh, xind, yejinc, jkautz, yidong}@nvidia.com

Abstract

Recent advances in reasoning-centric language models have highlighted reinforcement learning (RL) as a promising method for aligning models with verifiable rewards. However, it remains contentious whether RL truly expands a model's reasoning capabilities or merely amplifies high-reward outputs already latent in the base model's distribution, and whether continually scaling up RL compute reliably leads to improved reasoning performance. In this work, we challenge prevailing assumptions by demonstrating that prolonged RL (**ProRL**) training can uncover novel reasoning strategies that are inaccessible to base models, even under extensive sampling. We introduce ProRL, a novel training methodology that incorporates KL divergence control, reference policy resetting, and a diverse suite of tasks. Our empirical analysis reveals that RL-trained models consistently outperform base models across a wide range of pass@k evaluations, including scenarios where base models fail entirely regardless of the number of attempts. We further show that reasoning boundary improvements correlates strongly with task competence of base model and training duration, suggesting that RL can explore and populate new regions of solution space over time. These findings offer new insights into the conditions under which RL meaningfully expands reasoning boundaries in language models and establish a foundation for future work on long-horizon RL for reasoning. We release model weights to support further research:
https://huggingface.co/nvidia/Nemotron-Research-Reasoning-Qwen-1.5B

1 Introduction

Recent advances in reasoning-focused language models, exemplified by OpenAI-O1 [1] and DeepSeek-R1 [2], have marked a paradigm shift in artificial intelligence by scaling test-time computation. Specifically, test-time scaling enables long-form Chain-of-Thought (CoT) thinking and induces sophisticated reasoning behaviors, leading to remarkable improvements on complex tasks such as mathematical problem solving [3–6] and code generation [7, 8]. By continuously expending compute throughout the reasoning process—via exploration, verification, and backtracking—models boost their performance at the cost of generating longer reasoning traces.

At the heart of these advances lies reinforcement learning (RL), which has become instrumental in developing sophisticated reasoning capabilities. By optimizing against verifiable objective rewards rather than learned reward models, RL-based systems can mitigate the pitfalls of reward hacking [9–11] and align more closely with correct reasoning processes. However, a fundamental question remains under active debate within the research community: *Does reinforcement learning truly*

Preprint. Under review.

《PRORL：長期強化學習
拓展大型語言模型的推理邊界》
NVIDIA 2025.5.30

「菜，就多練」