Chain-of-Thought Prompting in Large Language Models

By Arsalan A. Khan | Portfolio: timedilationv2

Chain-of-Thought Prompting in Large Language Models

By Arsalan A. Khan | Portfolio: timedilationv2

Introduction

Chain-of-Thought (CoT) prompting is a technique that guides large language models (LLMs) to generate intermediate reasoning steps ("thoughts") before giving a final answer. This approach is especially effective for problems requiring multi-step reasoning, such as arithmetic, symbolic logic, and commonsense inference. CoT prompting requires no additional training-it only changes the format of the prompt by adding worked examples or reasoning instructions.

How It Works

Standard prompting gives the model a question and expects a direct answer. CoT prompting includes exemplars with reasoning steps, prompting the model to think step-by-step. For example, instead of "What is 12 divided by 0.5 plus 7?", CoT prompting produces: "12 divided by 0.5 is 24. Adding 7 gives 31." This stepwise logic significantly improves answer accuracy.

Performance Gains

- PaLM-540B: GSM8K accuracy rose from 55% (standard prompt) to 74% using CoT with self-consistency (+19%).

- SVAMP and AQuA tasks: +11% and +12% accuracy gains respectively.

- StrategyQA and CommonsenseQA: modest but consistent improvements.

These gains are especially pronounced in models with over 100B parameters. Smaller models often

generate incoherent reasoning chains, degrading accuracy.

Emergence at Scale

CoT benefits are seen predominantly in LLMs larger than 100B parameters. Below that scale, models tend to hallucinate or produce logical errors when asked to "think aloud." In contrast, large models show an emergent ability to reason step-by-step when prompted correctly.

Variants and Enhancements

- Zero-Shot CoT: A prompt like "Let's think step by step" alone can invoke reasoning without few-shot examples.
- Self-Consistency: Sampling multiple CoT paths and choosing the most common answer boosts accuracy significantly.
- Hybrid Systems: CoT has been combined with code interpreters, symbolic tools, and retrieval modules for enhanced performance.

Applications

CoT is effective in:

- Math word problems
- Commonsense reasoning
- Multi-hop QA
- Code reasoning
- Scientific and symbolic tasks

It is a core prompting strategy in systems like ChatGPT, Claude, and Gemini.

Conclusion

Chain-of-Thought prompting is a foundational method in the evolution of prompt engineering. By eliciting intermediate reasoning from LLMs, it transforms their ability to solve complex tasks without retraining. As model scale increases, CoT unlocks new emergent reasoning capabilities, highlighting the profound link between model size, prompt format, and intelligence expression.