Chain-of-Thought Prompting: A Review

By Arsalan A. Khan | Portfolio: timedilationv2

Chain-of-Thought (CoT) prompting is a technique that improves large language models (LLMs) by encouraging them to "think step-by-step" when solving complex problems.

Key Performance Gains:

- PaLM-540B saw GSM8K benchmark accuracy rise from 55% (standard) to 74% with CoT + self-consistency (+19% absolute gain).

- Double-digit improvements observed in SVAMP (+11%), AQuA (+12%), and StrategyQA.

- CoT works best on models >100B parameters; smaller models often perform worse with CoT prompting.

Why It Works:

- CoT unlocks latent reasoning by prompting LLMs to explain intermediate steps.

- Helps reduce shallow guesses, improves interpretability.

Variants & Enhancements:

- Zero-Shot CoT: Prompts like "Let's think step by step" trigger reasoning without few-shot examples.

- Self-Consistency Decoding: Samples multiple reasoning paths and chooses the most frequent answer - leading to robust performance improvements.

Use Cases:

- Arithmetic word problems, symbolic logic, commonsense reasoning, multi-hop QA, programming,

and planning.

- Forms the backbone of advanced systems like ChatGPT, Claude, and Gemini.

Implications:

- No model retraining required - only prompt changes.

- Demonstrates that large models develop emergent reasoning capabilities.

- Core component in advanced prompt engineering workflows.

Summary:

Chain-of-Thought prompting represents a major breakthrough in prompting strategies. It enables more logical, interpretable, and accurate outputs from large language models, often surpassing fine-tuned baselines by simply providing structured reasoning in prompts.