Filler tokens and problem repeats increase no Chain-of-Thought (CoT) math performance

Experimental Setup

Baseline (No CoT)

Problem: How many 8-digit integers become 1/9 as big if the first digit is removed?

→ *LLM* → Answer: 8

Filler Tokens

Problem: How many 8-digit...

Filler: 1 2 3 ... 300

→ *LLM* → Answer: 7

Repeats

Problem: How many 8-digit...
Problem (repeat 2): ...
Problem (repeat 3): ...

→ *LLM* → Answer: 7

Legend:
- Filler tokens (unrelated text)
- Problem repetitions

Opus 4.5 on Easy-Comp-Math

Bar chart showing Accuracy (%):
- Baseline (No CoT): 45.2%
- Filler tokens (f=300): 51.0%
- Repeats (r=5): 51.2%

$p < 0.001$ (Baseline vs Filler tokens)
$p < 0.001$ (Baseline vs Repeats)