

# The Practicality of Prompt Engineering

- ▶ Cost (length/complexity, financial) efficacy of prompt engineering
- ▶ 100 programmatic conversations - math (GSM8K), creative writing (CW here)
- ▶ Contribution - quantification, statistical tests, old + new models
- ▶ Below are results on gains versus no prompt engineering
- ▶ Statistically significant improvements (paired tests): GSM8K all, creative writing GPT-4 manual CoT (Chain of Thought), zero-shot CoT
- ▶ Zero-Shot CoT cheaper, more powerful than prior and later techniques
- ▶ TD3 (Nov 2022) to GPT-4 (Jun 2023)- GSM8K gains down, CW gains up
- ▶ A few pennies per 1K tokens - cheap GSM8K, expensive CW improvements

