

# The Practicality of Prompt Engineering

- ▶ Cost (length/complexity, financial) efficacy of prompt engineering
- ▶ 100 programmatically-conducted conversations
- ▶ Contribution - quantification, statistical tests, old + new models
- ▶ Statistically significant improvements (paired tests): GSM8K all, creative writing GPT-4 manual CoT, zero-shot CoT
- ▶ Zero-Shot CoT cheaper, more powerful than few-shotting that came before, complicated techniques that came after
- ▶ TD3 to GPT-4 - GSM8K gains down, CW gains up
- ▶ A few pennies per 1K tokens - cheap GSM8K, expensive CW improvements

