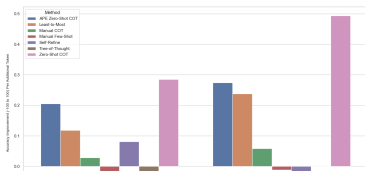
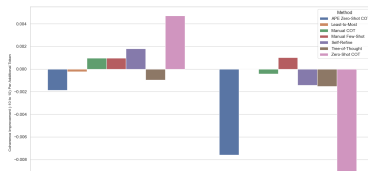


The Practicality of Prompt Engineering

- ▶ What prompt engineering methods are most effective for a given cost (financial, length, complexity)?
- ▶ Data: 100 programmatically-conducted conversations
- ▶ Novel contribution - new metrics, significance testing, run on new and old models
- ▶ Statistically Significant Improvements (Paired Tests): GSM8K All, Creative Writing GPT-4 Manual CoT, Zero-Shot CoT
- ▶ Gains on CW are growing but gains on GSM8K are shrinking in transition from TD3 to GPT-4
- ▶ Gains growing/shrinking on gpt4, cost of tokens and human review, newer techniques more effective, variance



Isaac Liu



The Practicality of Prompt Engineering