GPT-4 Performance, All Modalities, All Datasets

Legend:
- Zero Shot
- Zero Shot CoT
- Suppressed CoT
- Explanation First
- Answer First

MultiArith
- Zero Shot: 98%
- Zero Shot CoT: 98%
- Suppressed CoT: 90%
- Explanation First: 98%
- Answer First: 94%

GSM8k
- Zero Shot: 93%
- Zero Shot CoT: 95%
- Suppressed CoT: 43%
- Explanation First: 95%
- Answer First: 55%

Aqua-RAT
- Zero Shot: 52%
- Zero Shot CoT: 57%
- Suppressed CoT: 36%
- Explanation First: 68%
- Answer First: 45%

Coin Flip
- Zero Shot: 56%
- Zero Shot CoT: 100%
- Suppressed CoT: 58%
- Explanation First: 100%
- Answer First: 57%

MMLU
- Zero Shot: 57%
- Zero Shot CoT: 62%
- Suppressed CoT: 48%
- Explanation First: 60%
- Answer First: 48%