GPT-4 Performance on Stepwise Dataset

