

Conversation History Task

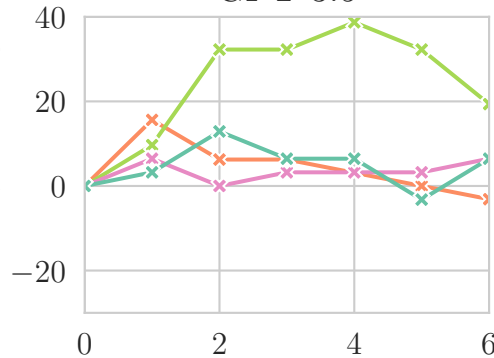
—x— Gigaword

—x— MMLU Abstract Algebra

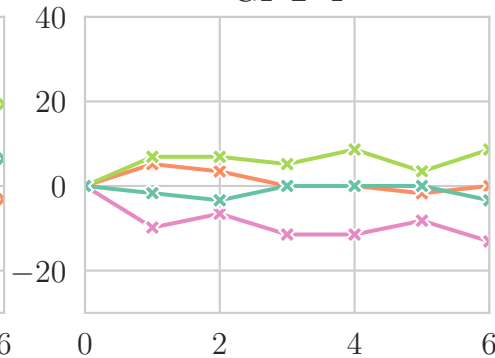
—x— Rotten Tomatoes

—x— TweetQA

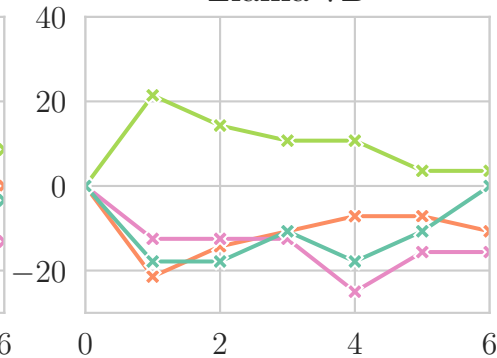
GPT-3.5



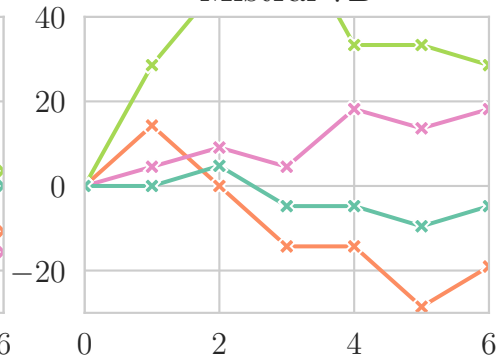
GPT-4



Llama-7B



Mistral-7B



Conversation History Length L