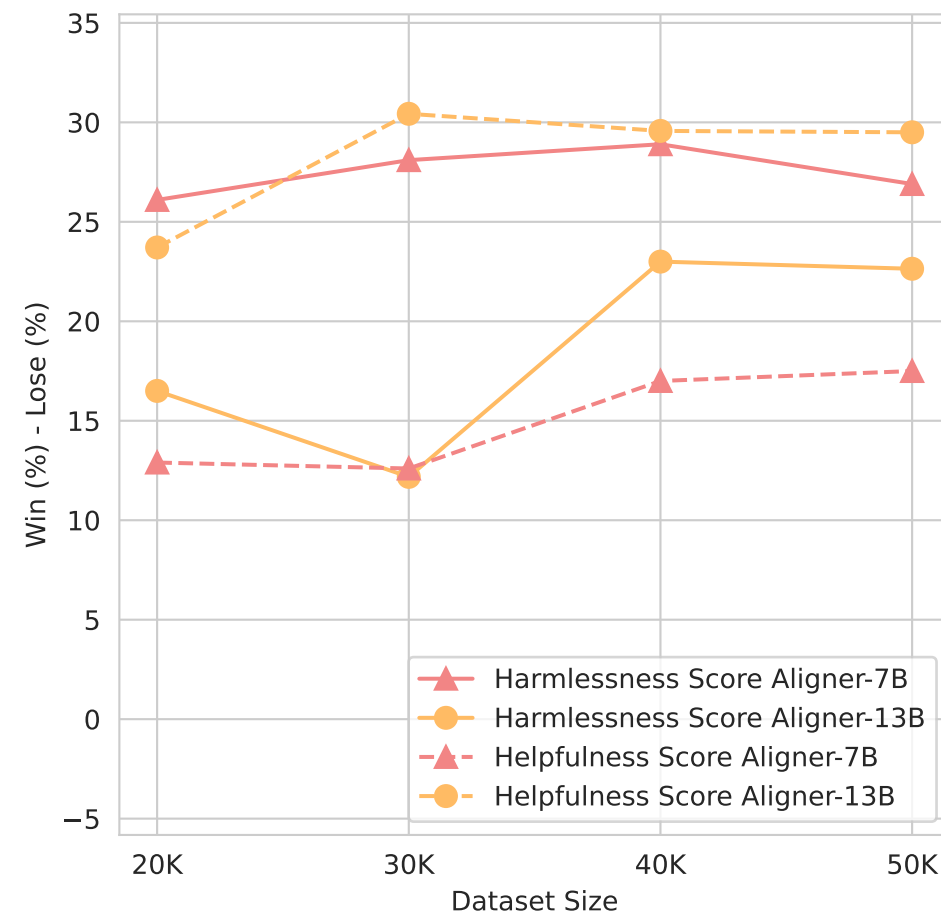
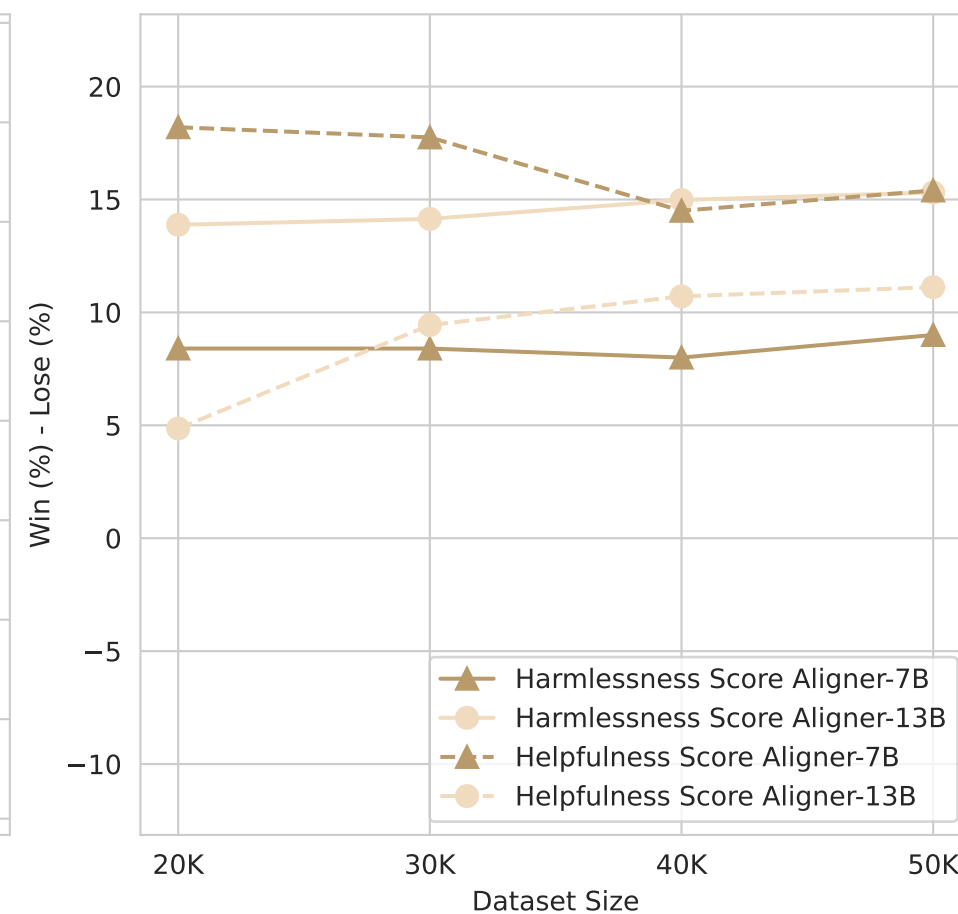


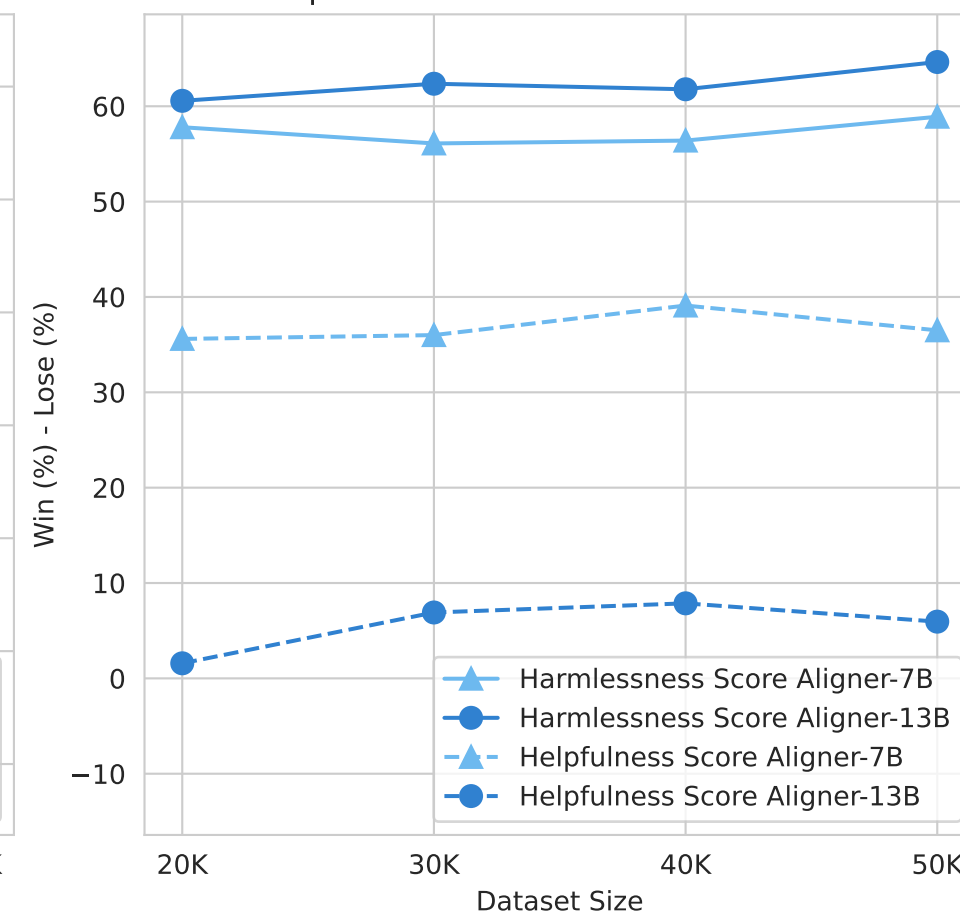
GPT-4: Performance Evaluation



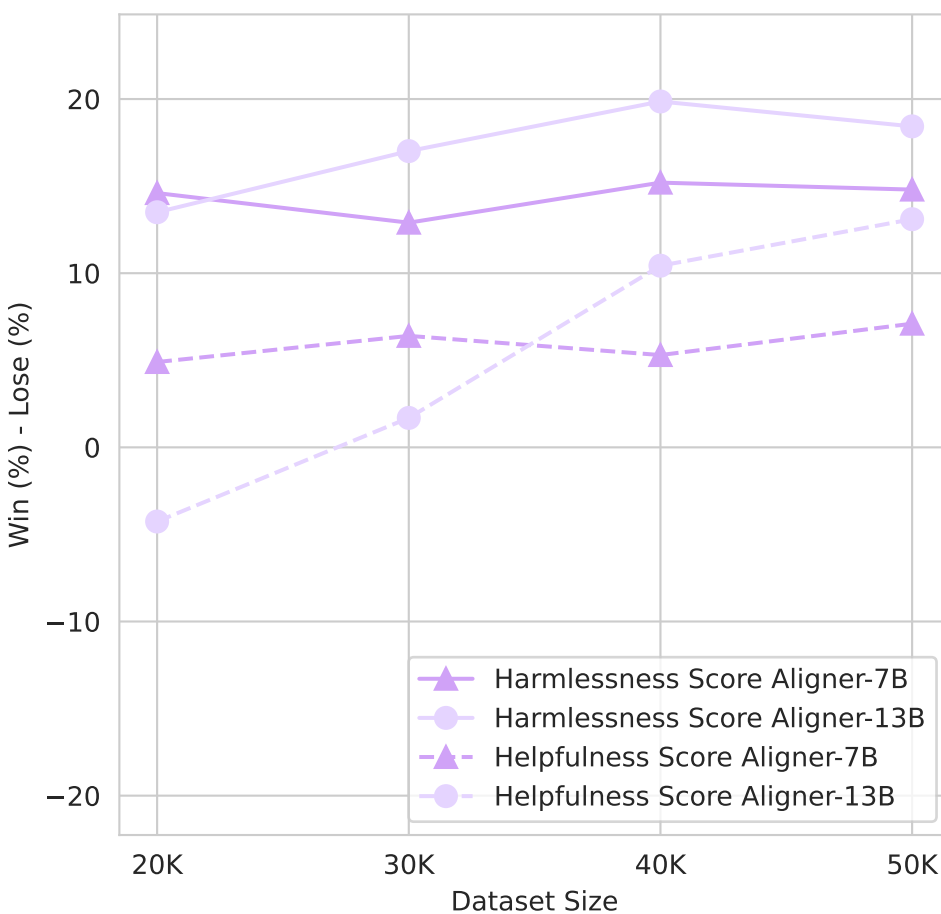
Beaver-v1: Performance Evaluation



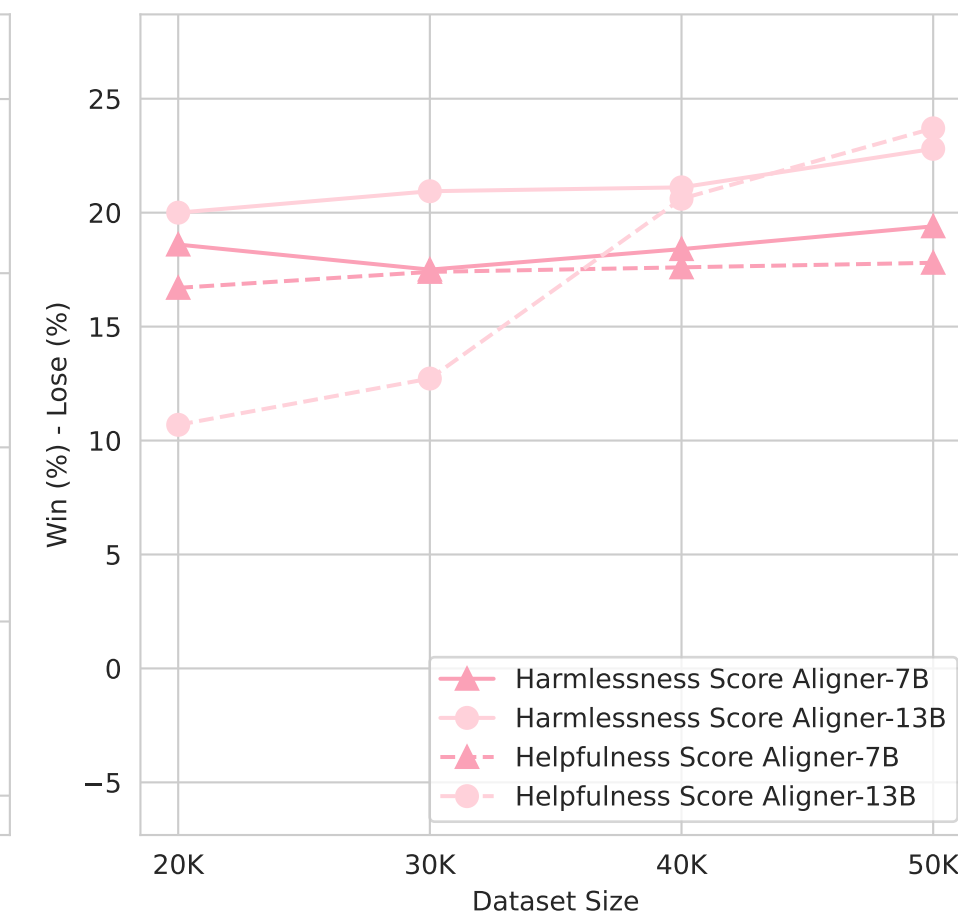
Alpaca-7B: Performance Evaluation



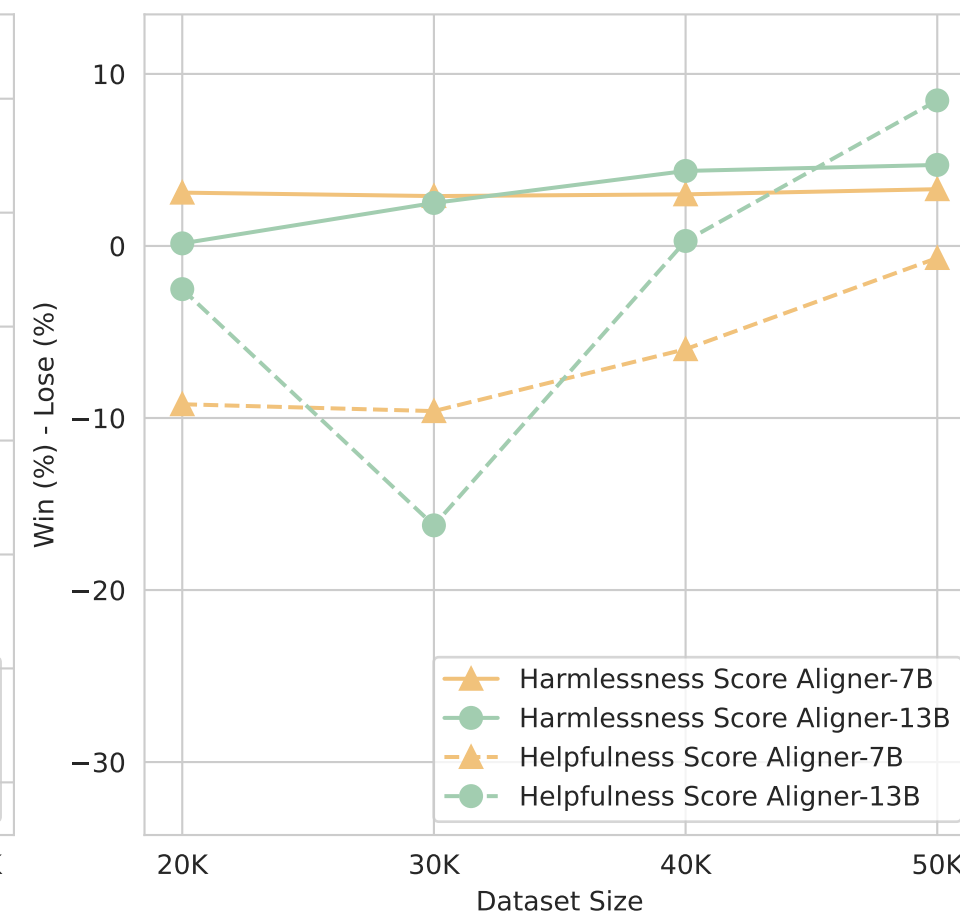
Llama 2-7B-Chat: Performance Evaluation



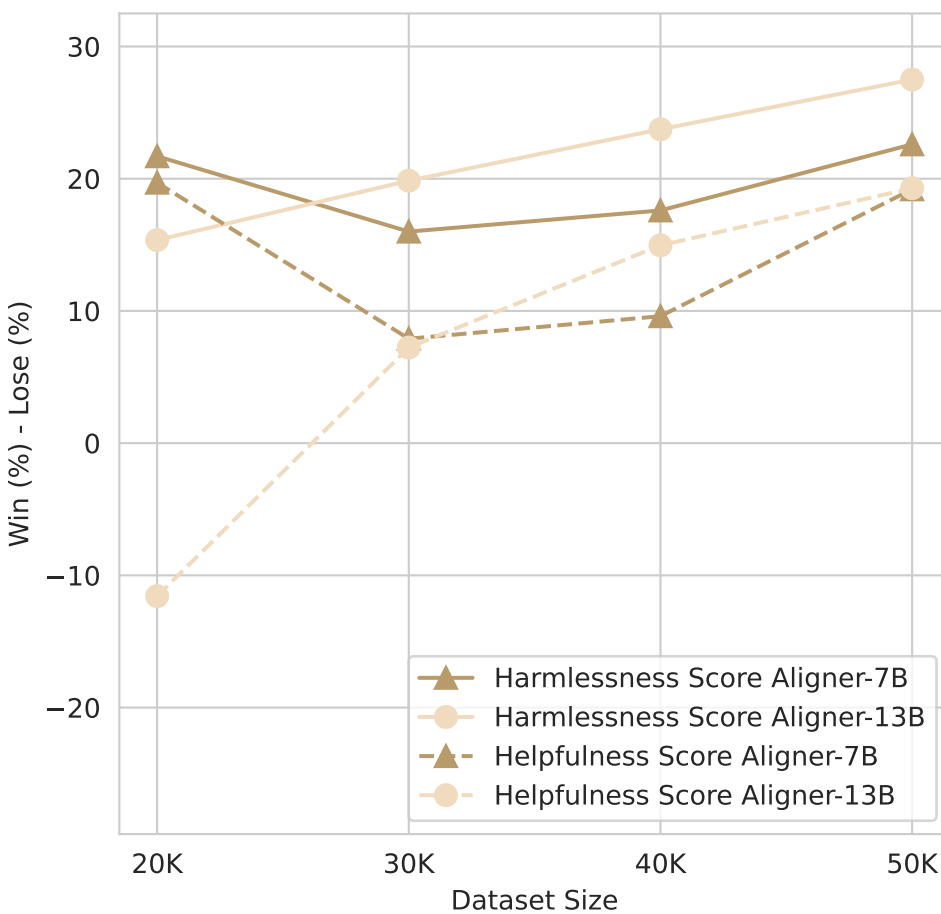
Llama 2-13B-Chat: Performance Evaluation



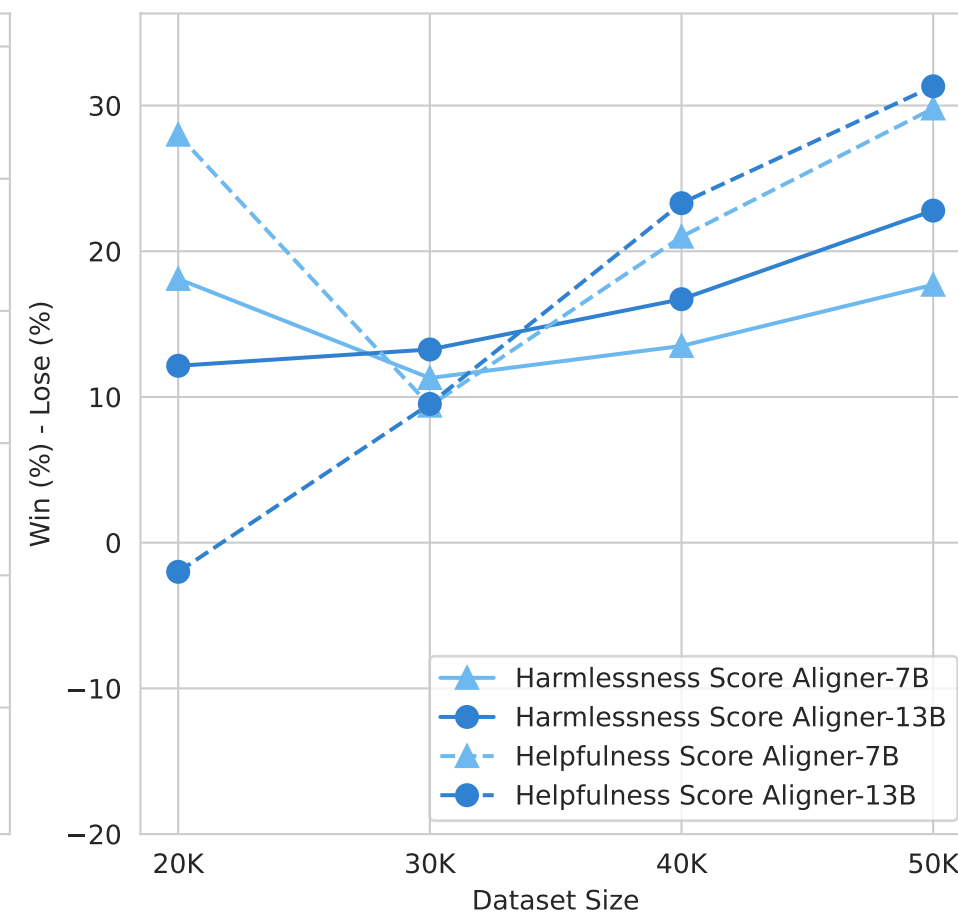
Llama 2-70B-Chat: Performance Evaluation



Vicuna-7B: Performance Evaluation



Vicuna-13B: Performance Evaluation



Vicuna-33B: Performance Evaluation

