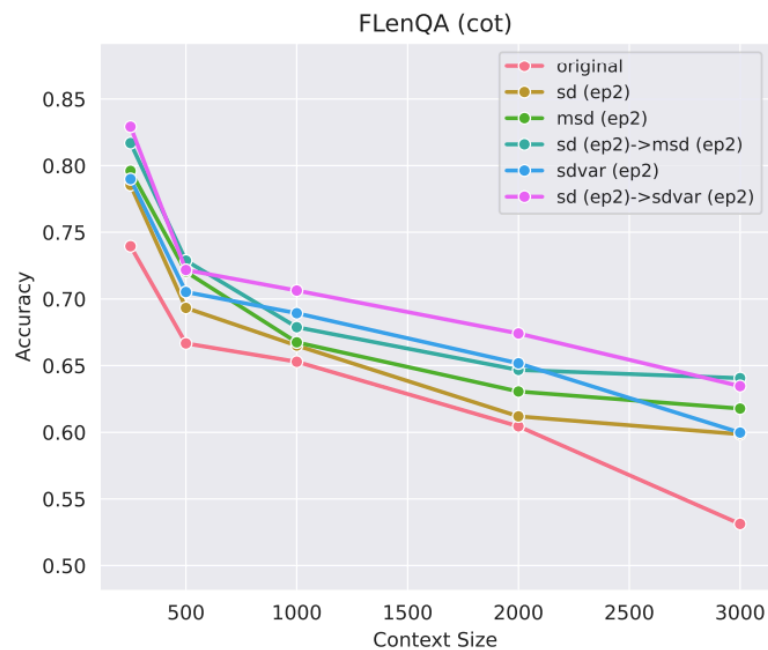
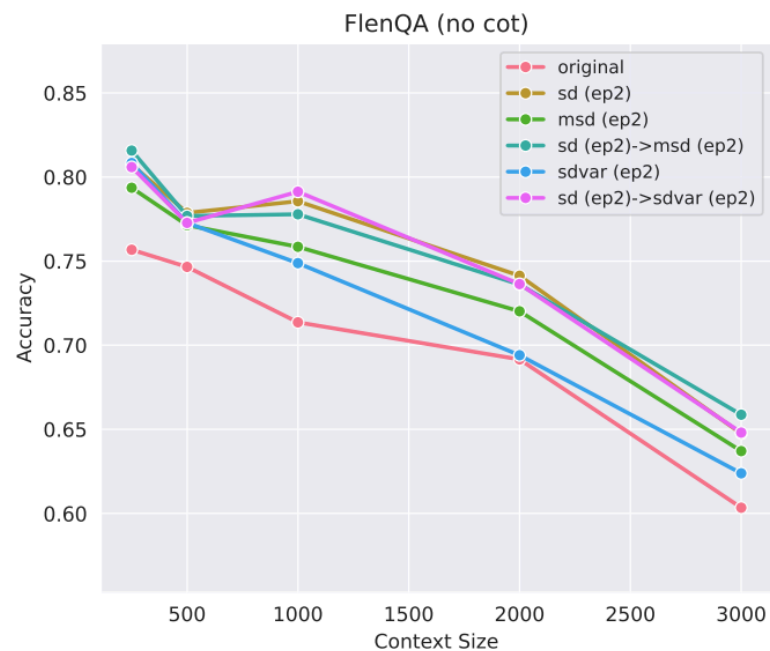


(a) MDQA



(b) FLenQA with chain-of-thought prompting



(c) FLenQA without chain-of-thought prompting

Figure 16: Performance of finetuned Mistral 7B with different retrieval tasks.