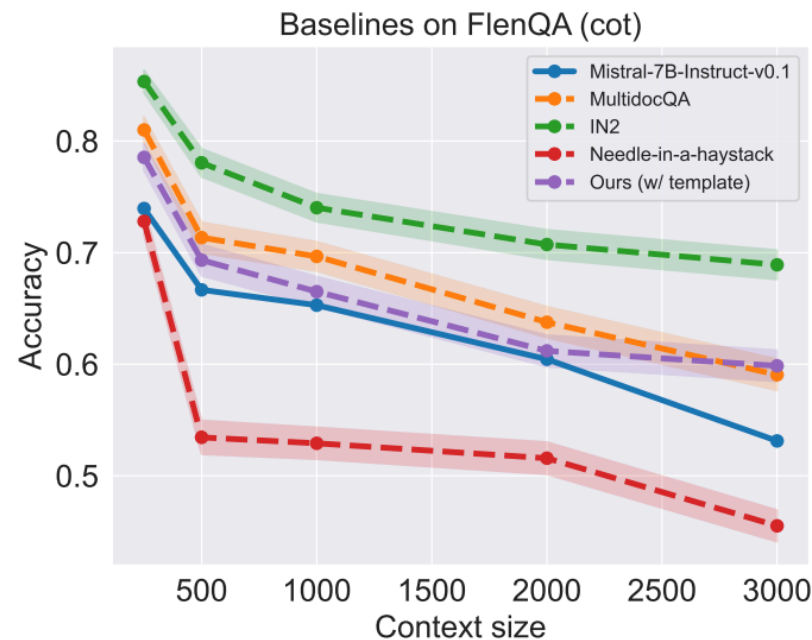
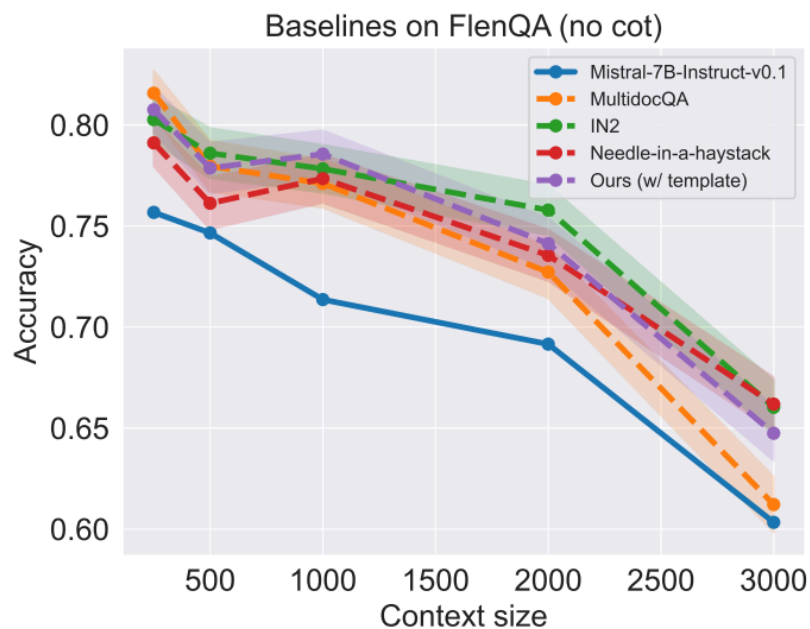


(a) MDQA



(b) FLenQA with chain-of-thought prompting



(c) FLenQA without chain-of-thought prompting

Figure 8: Performance of finetuned Mistral 7B on (a) MDQA, (b) FLenQA with chain-of-thought prompting, and (c) FLenQA without chain-of-thought prompting.