

<b>Finetuning dataset</b>	<b>MMLU</b>	<b>HellaSwag</b>	<b>GSM8K</b>	<b>TriviaQA</b>	<b>NQ-Open</b>
Mistral-7B-Instruct-v0.2	58.74	65.98	41.85	32.65	14.46
Ours (w/ template)	58.42 (−0.32)	65.74 (−0.24)	41.72 (−0.13)	33.02 (+0.37)	15.03 (+0.57)

Table I: The performance of Mistral-7B-Instruct-v0.2 and finetuned version (on simple dictionary key-value retrieval with context length of 24K) on general capability benchmarks. All numbers are reported in percentage.