Finetuning dataset	MMLU	HellaSwag	GSM8K	TriviaQA	NQ-Open
Original	53.42	56.31	34.65	47.63	11.61
sd (ep1)	53.38 ($-$0.04)	56.26 ($-$0.05)	34.58 ($-$0.07)	47.54 ($-$0.09)	11.97 (+0.36)
sd (ep2)	53.44 (+0.02)	56.22 ($-$0.09)	34.34 ($-$0.31)	47.74 (+0.11)	11.98 (+0.37)
sd (ep4)	53.29 ($-$0.13)	56.20 ($-$0.11)	34.19 ($-$0.46)	47.63 (+0.00)	11.85 (+0.25)
sd x2 (ep2)	53.35 ($-$0.07)	56.30 ($-$0.01)	33.89 ($-$0.76)	47.83 (+0.20)	11.95 (+0.34)

Table 3: Mistral 7B and finetuned versions' performance evaluated on general ability benchmarks. All numbers are reported in percentage.