Finetuning dataset	MMLU	HellaSwag	GSM8K	TriviaQA	NQ-Open
Original	53.42	56.31	34.65	47.63	11.61
msd (ep2)	53.36 (−0.06)	56.29 (−0.02)	34.31 (−0.34)	47.81 (+0.18)	11.84 (+0.23)
sd (ep2)→msd (ep2)	53.28 (−0.14)	56.21 (−0.10)	33.78 (−0.87)	47.81 (+0.18)	11.82 (+0.21)
sdvar (ep2)	53.39 (−0.03)	56.26 (−0.05)	34.28 (−0.37)	47.66 (+0.03)	11.81 (+0.20)
sd (ep2)→sdvar (ep2)	53.16 (−0.26)	56.15 (−0.16)	33.72 (−0.93)	47.60 (−0.03)	11.89 (+0.28)

Table 6: Mistral 7B and finetuned versions' performance evaluated on general ability benchmarks. All numbers are reported in percentage.