Finetuning dataset	MMLU	HellaSwag	GSM8K	TriviaQA	NQ-Open
Original	53.42	56.31	34.65	47.63	11.61
IN2 (ep1)	53.27 (−0.15)	56.26 (−0.05)	34.65 (+0.00)	45.59 (−2.03)	10.00 (−1.61)
IN2 (ep2)	53.49 (+0.07)	56.44 (+0.13)	34.98 (+0.32)	45.44 (−2.19)	9.80 (−1.81)
IN2 (ep4)	53.37 (−0.05)	56.69 (+0.38)	34.91 (+0.26)	43.98 (−3.65)	7.47 (−4.14)
IN2 x2 (ep2)	53.31 (−0.11)	56.68 (+0.37)	33.89 (−0.76)	44.80 (−2.83)	9.43 (−2.18)

Table 5: Mistral 7B and finetuned versions' performance evaluated on general ability benchmarks. All numbers are reported in percentage.