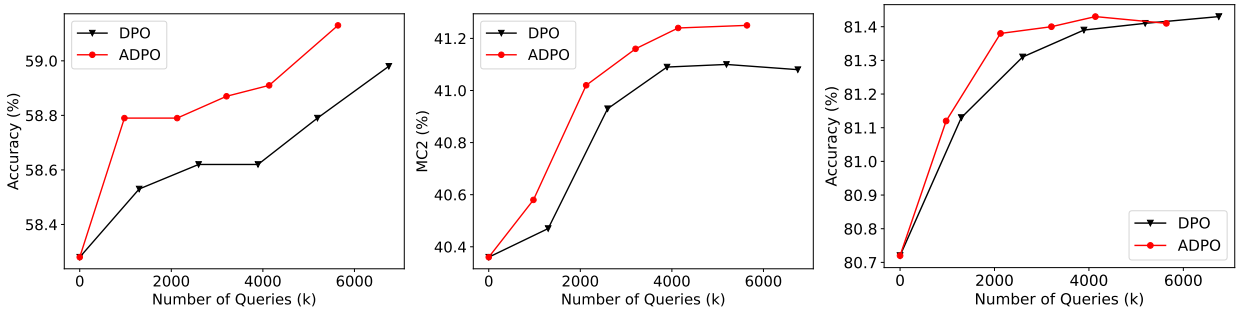
Table 1: Results of Zephyr-Beta trained on dpo-mix-7k dataset on objective benchmarks. ADPO outperforms DPO on ARC, and performs comparably to DPO on, TruthfulQA and HellaSwag, resulting higher average performances. Besides, ADPO only makes 5.6k queries, which is fewer than the queries made by DPO.

Models	ARC	TruthfulQA	HellaSwag	Average	# Queries
Zephyr-Beta-SFT	58.28	40.36	80.72	59.79	0
Zephyr-Beta-DPO	58.53	41.08	**81.43**	60.35	6751
Zephyr-Beta-ADPO	**59.13**	**41.25**	81.41	**60.61**	**5640**

Table 2: Results of Zephyr-Beta trained on dpo-mix-7k dataset on on subjective benchmarks including AlpacaEval 2.0 and MT-Bench. Here WR stands for win rate and LC stands for length controlled. ADPO also achieves comparable performance with DPO.

Models	MT-Bench			Alpaca Eval 2.0		
	First Turn	Second Turn	Average	LC WR	WR	Avg. Length
Zephyr-Beta-SFT	**6.82**	5.94	6.39	4.59	4.69	1741
Zephyr-Beta-DPO	6.71	6.39	6.55	6.67	**5.63**	1438
Zephyr-Beta-ADPO	6.67	**6.47**	**6.57**	**6.88**	5.57	1392



(a) ARC Challenge (b) TruthfulQA (c) HellaSwag

Figure 1: The test accuracy curve of DPO and ADPO starting from Zephyr-Beta-SFT training on dpo-mix-7k. The x-axis is the number of queries and the y-axis is the metric for corresponding dataset. Compared to DPO, ADPO enjoys a faster performance improvement and a higher performance upper bound.