Table 1: Mean Values for Different Methods of the final checkpoint

	SFT	DPO-Shift $f(\lambda) = 0.95$	DPO	IPO	SimPO
Mean of final chosen logp	-299.09	-324.86	-418.80	-591.85	-404.46
Mean of final rejected logp	-278.60	-356.67	-451.08	-644.37	-412.29
Mean of final reward margin	N/A	0.52	0.53	0.73	0.28