

DPO / CITA: Reward Accuracy (Llama-3.1-8B)

