

DPO vs CITA: Reward Margins (Llama-3.1-8B)

