

Direct Preference Optimization (DPO)

What is DPO?

DPO directly optimizes the language model using preference data, eliminating the need for a separate reward model and RL training loop.

DPO vs PPO: Architecture Comparison

