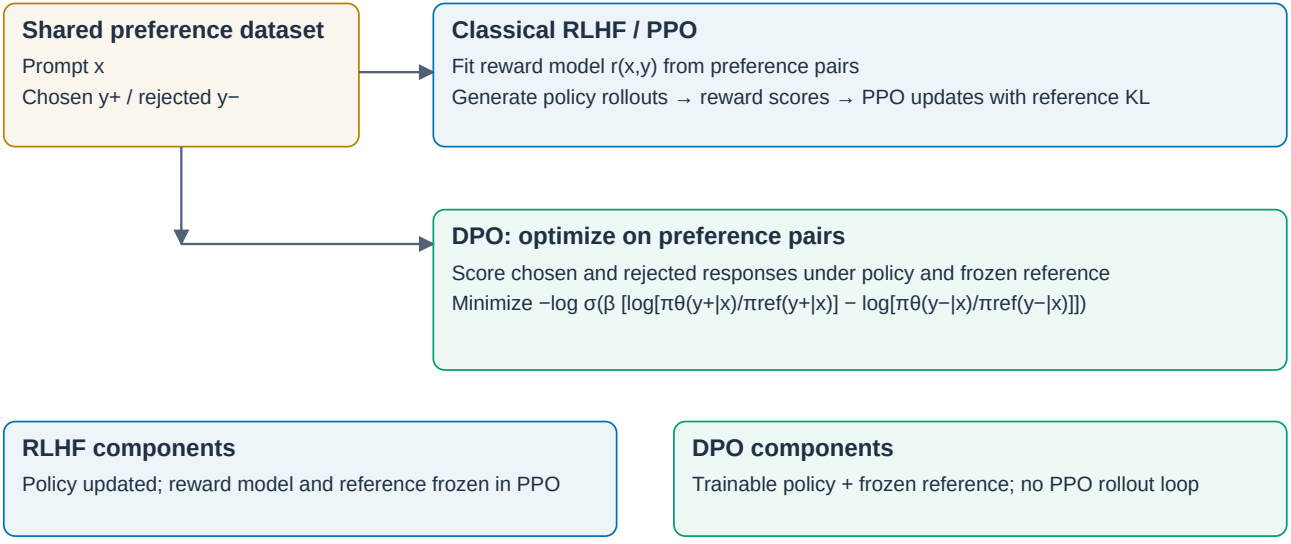
Preference alignment: RLHF / PPO versus DPO

A training-path comparison; both paths begin with preference data and an initial policy.



Original educational redraw · schematic values only · SVG text and objects remain editable