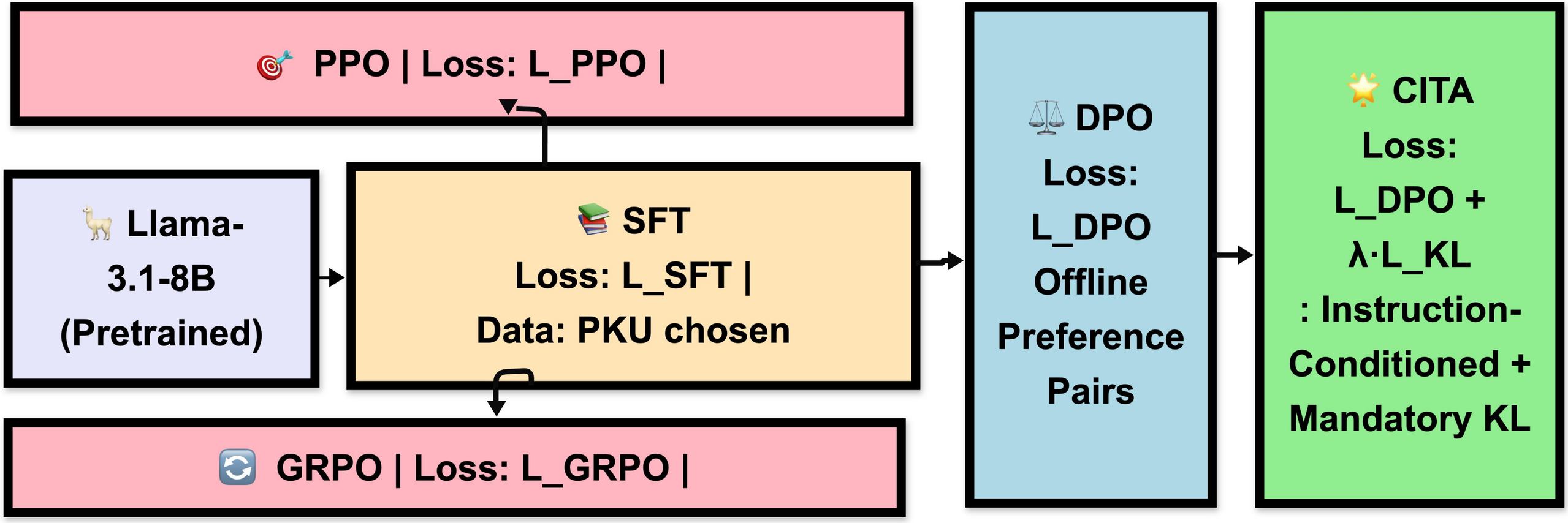
🎯 PPO | Loss: L_PPO |
🦙 Llama-3.1-8B (Pretrained)
📚 SFT
Loss: L_SFT |
Data: PKU chosen
🔄 GRPO | Loss: L_GRPO |
⚖️ DPO
Loss: L_DPO
Offline Preference Pairs
🌟 CITA
Loss: L_DPO + λ·L_KL
: Instruction-Conditioned + Mandatory KL