Online vs Offline RLHF

Offline RLHF

Train on fixed datasets of preferences collected in advance. More common in medical AI due to safety requirements.

Offline Advantages

- Safety: All data reviewed before training
- Reproducibility: Fixed dataset enables consistent experiments
- Expert Efficiency: Collect preferences in batches
- Lower Risk: No live patient interaction during training

Online RLHF

Continuously collect new preferences during deployment and retrain. Riskier but more adaptive to evolving needs.

Online Advantages

- Adaptation: Quickly respond to new medical knowledge