

Reinforcement Learning from Human Feedback

RLHF for Medical Domain

Align model outputs with medical expert preferences



10K+

Comparisons

PPO

Algorithm

+15%

Expert Preference

Medical Expert Feedback

- **Accuracy:** Factual correctness of medical information
- **Safety:** Avoidance of harmful recommendations
- **Clarity:** Patient-friendly explanations
- **Completeness:** Comprehensive coverage of topic