

# Preference Dataset Creation

## Dataset Structure

Preference datasets contain pairs of model outputs with expert rankings, forming the foundation for reward model training.

## Data Components



### 1. Input Prompt

*Clinical query or task description*



### 2. Output A

*"Patient presents with fever and cough. Recommend chest X-ray and CBC."*



### 3. Output B

*"Recommend immediate chest X-ray, CBC, and respiratory pathogen panel given symptoms."*



### 4. Preference Label



*Expert prefers B: More comprehensive diagnostic approach*



### 5. Confidence Score (Optional)

Strength: High (0.9) - Clear preference for comprehensive testing



## 6. Rationale

Expert's reasoning: "Output B includes respiratory pathogen panel which is essential for differential diagnosis in current clinical context."

## Dataset Statistics

**10K-100K+**

Preference Pairs

**50/50**

Balance (A vs B)

**15+ Specialties**

Clinical Coverage

**85-95%**

Agreement Rate