

# Blind Evaluation Setup



## Single-Blind

Evaluators unaware of model identity



## Double-Blind

Both evaluators and data collectors blinded

### Blind Evaluation Process



Randomize



Anonymize



Evaluate



Reveal & Analyze



#### Positive Controls

Known good responses



#### Negative Controls

Known poor responses



#### Baseline Comparison

Compare to human experts



#### Neutral Controls

Ambiguous cases

Minimize bias: Anonymize responses, randomize order, use control cases