

# Evaluation Framework Overview



## Automated Evaluation

Metrics computed automatically on benchmark datasets



## Human Evaluation

Expert clinicians assess clinical relevance and safety



## Medical-Specific

Domain-specific metrics for healthcare applications

## Key Evaluation Categories

- Accuracy & Factuality
- Safety Assessment
- Consistency Measures
- Clinical Relevance
- Hallucination Detection
- Bias & Fairness