

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 | • Clinical Relevance |
| • Safety Assessment | • Hallucination Detection |
| • Consistency Measures | • Bias & Fairness |