

## Part 2/3:

# Medical-Specific Metrics

- Clinical utility vs. accuracy tradeoffs
- Hallucination detection and factuality
- Consistency and uncertainty quantification
- Calibration and confidence scoring