

# Consistency Measures



Response over time



Paraphrase robustness



Multi-query consistency



Temperature sensitivity

## Testing Methodology

- 1 Ask same question 10+ times
- 2 Rephrase in different ways
- 3 Vary context and details
- 4 Measure agreement rate
- 5 Identify variation sources
- 6 Analyze error patterns

0%

50%

90%

100%



Target: >90% consistency for critical clinical decisions