

Agent Evaluation Metrics

Performance

- Accuracy
- Precision
- Recall
- F1-Score

Measures how well the agent completes tasks correctly, including prediction accuracy, precision in identifying relevant information, recall in capturing all necessary elements, and balanced F1-score.

Efficiency

- Response Time
- Throughput
- Resource Usage
- Latency

Evaluates operational speed and resource optimization, tracking response time, task throughput, computational resource consumption, and system latency for optimal performance.

Reliability

- Uptime
- Error Rate
- Recovery Time
- Fault Tolerance

Assesses system stability and dependability through uptime monitoring, error frequency analysis, recovery speed after failures, and capacity to handle unexpected conditions.

Safety

- Error Detection
- False Alarm Rate
- Clinical Impact
- Patient Safety

Ensures safe operation in critical environments by monitoring error detection capabilities, minimizing false alarms, evaluating clinical outcomes, and prioritizing patient safety standards.