

# Dimension

## Focus

# AI observability

System health, infrastructure, cost, model performance metrics

## Failure types detected

Latency spikes, model drift, token waste, infrastructure outages, cost overruns

## Data needed

Logs, metrics, traces, cost reports, infrastructure telemetry

## Typical tools

Monitoring dashboards, APM systems, cost trackers, alerting platforms

## Primary question

"Is the system running correctly?"

## Strongest value

Operational reliability and cost efficiency

# Semantic observability

Output meaning, reasoning quality, intent alignment, factual accuracy

Hallucinations, reasoning errors, intent misalignment, bias in generated content, retrieval quality degradation

Reasoning traces, evaluation scores, retrieval quality assessments, human feedback signals

Evaluation frameworks, LLM-as-judge scoring, human review workflows, reasoning trace analyzers

"Is the system producing correct and appropriate outputs?"

Trust, compliance, and output quality assurance