

Metric

What it measures

Target

Measurement approach

Exactness

Percentage of numeric results matching expected values within tolerance

99%

SQL validation against golden dataset

Faithfulness

Percentage of claims traceable to verifiable source data

100%

Pass/fail per assertion; automated lineage check

Coverage

Ratio of narrative content supported by specific citations or data points

70%+

Word-level annotation of cited vs. uncited content

Governance conformance

Percentage of queries resolved using certified semantic layer definitions

95%+

Query metadata tagging against certified entity registry

P95 latency

95th percentile response time

<2s (lookup), <10s (analysis)

Request-level timing instrumentation

Cost efficiency

Average token cost per 100 queries

Budget-dependent

Token usage aggregation per query batch