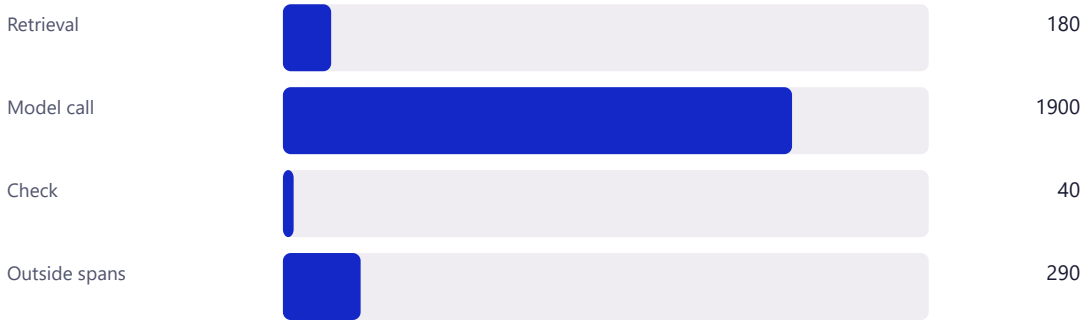
CHAPTER 20 / OBSERVABILITY

Healthy API. Wrong answer.

Compare what a trace explains with what it exposes.

Compare what a trace explains with what it exposes.



Handpicked example values · inspect exact values below

Illustrative monitoring layers. Thresholds depend on your service objectives; minimize or redact sensitive logging.

Infrastructure

Track availability, errors, queueing and latency.

Request

Track token usage, version, time to first token and cost estimates.

Outcome

Use review and task-specific evaluations to spot unsupported or unhelpful answers.

↗ **Track availability, errors, queueing and latency.**

PUT IT TO WORK

An API can return HTTP 200 while quoting the wrong policy. Request traces help locate it; quality review detects the wrong claim.

CLOSE THE PAGE. EXPLAIN IT.

Which signal would reveal a fluent but unsupported answer when uptime stays green?