

ILLUSTRATIVE DELIVERABLE / SUPPLIED RUN DATA

Supplier acceptance. Two required checks failed.

BLOCKERS

2 required checks failed

route	Intermediary A · chat/completions
model requested	gpt-4o
mode	standard · 321 requests
profile	platform-team.v1
spend	\$0.044 of \$0.50 cap

Decision context

The reference test did not reject the route, but the route failed two mandatory contract checks: max_tokens and unknown_parameter. Reference consistency does not override those failures.

Provenance and limits

This PDF reproduces the supplied example run. It is not a new audit and has no cryptographic signature. The original example labels its report "sha256 signed"; SHA-256 by itself is a hash, not a digital signature. The measured route is anonymised.

01 / API CONTRACT

What the route honours.

availability PASS

Endpoint reachable, TLS valid, no redirects followed.

streaming PASS

Real SSE. A 230-character probe arrived in 14 content deltas — not a buffered response replayed as a stream.

tool_loop PASS

Tool call issued, result accepted, final message produced.

json_mode PASS

Valid JSON returned on every attempt.

structured_output WARN

Schema honoured, but an optional field was emitted that the schema does not declare.

max_tokens FAIL

Limit set to 128. Completions returned 681–942 tokens on all five models tested. Reproduced across three runs.

unknown_parameter FAIL

A deliberately misspelled, wrongly typed parameter was accepted with 200 and silently discarded.

error_handling WARN

Unknown model name returns 503 rather than a 4xx. Marked optional in this profile.

usage_sanity WARN

Usage object carries one field outside the documented schema.

Consistency has a scope.

exact_reply PASS

Instruction followed literally at an 8-token budget.

json_shape PASS

Required keys present, types correct.

tool_selection PASS

Correct tool chosen over the decoy.

refusal_boundary PASS

Refused as expected; no policy leakage into the response body.

long_context NOT TESTED

Outside this profile's declared requirements.

REFERENCE CONSISTENCY

outcome NOT REJECTED

The route was not rejected against the approved reference — but comparison strength is reduced, because the model identifier is self-reported by the provider.

e-value peak 1.43

Rejection threshold is 100. The sequential test is anytime-valid: it may stop early, and it did not.

aggregate distance 0.143

Jensen–Shannon divergence over 300 paired samples, zero errors.

reference tier A

Enrolled directly from the first-party vendor, generation 2, 411 observations, validity 1.00.

03 / PERFORMANCE, COST AND HISTORY

A record you can compare.

latency median **2.09 s**

Measured from the machine that ran the test — a client-side number, not the provider's.

latency p95 **8.86 s**

Long tail concentrated in the fingerprint battery, n=300.

time to first chunk **3.0 s**

Streaming path.

requests **321**

Every attempt logged, including retries.

cost **\$0.044**

Priced from the profile's own rate table, not the provider's invoice.

reported ids **2 distinct**

316 responses carried one snapshot identifier, 3 carried an older one. The route occasionally answers from a different snapshot.

Previous comparable run

Compared with the previous run of this profile: reference consistency moved from INCONCLUSIVE to NOT REJECTED under the same conditions. Two contract checks are unchanged failures.