

Context-Control Security and Data Boundary

An evidence summary for a technical or procurement reviewer. Every statement below is traceable to committed source, a test, or a published artifact, and the mapping is machine-checkable. It claims no certification, no compliance status, no partnership, and no guaranteed outcome. Where a boundary is narrower than it might sound, the narrow version is the one written down.

PURPOSE AND SCOPE

- This covers two things: the Maha Context Compiler, and the bounded interceptor that runs it inside a WSO2 AI Gateway request.
- It does not describe your gateway, your model provider, your cloud account, or your deployment. Those keep their own retention, logging and network controls, and this document makes no statement about them.
- It is an evidence summary, not a substitute for your own security review.

WHAT IS HANDLED

- A request carries a task string, a token budget, and one to eight documents, each with an identifier, an optional title, and text.
- The response returns the compiled context pack, the selected passages verbatim with their identifiers and hashes, a per-source manifest, token and coverage metrics, explicit warnings, and input and output hashes.
- On the metered direct API, the only record kept of a call is a usage row of five fields: access mode, credential identifier, status class, and the compiler's own input and output token estimates. No task, no document, no identifier from your content, and no hash is written to it.

THE SOURCE-TEXT BOUNDARY

Stated precisely, because the useful version of this claim is narrower than the marketing version.

- Your source text is processed in the request that carries it. Selection is deterministic ranking and de-duplication over that text; no model is invoked and nothing is sent to a third party by the compiler.
- The compiler returns the passages it selected, verbatim. That is the product, not a leak: the pack is the text you asked it to choose from your own documents, and it is returned to the caller who supplied them.
- In the WSO2 interceptor path, neither route handler imports or invokes any database, cache, queue or filesystem client, and both mark their responses no-store. What the gateway receives back is metadata: a pack identifier, hashes, counts, token measurements and a bypass decision.
- Where evidence is retained, it is hashes and metadata: SHA-256 over each normalised source, a hash per selected passage, source and passage identifiers, passage counts, coverage percentages and token estimates.
- Published and sanitised artifacts exclude document bodies, compiled context, request headers and credentials, and declare those exclusions in the artifact itself.
- Maha does not claim universal zero retention. The claim above is scoped to these components. Your gateway, your provider and your own logging may retain the same text under settings Maha neither sets nor sees.

INTEGRITY AND EVIDENCE

- The input hash commits to the task, the token budget, and for each document its identifier, title and a SHA-256 of its normalised text. Document bodies are not part of the hash preimage.
- The output hash is a SHA-256 over the exact rendered pack, so the bytes a model received can be identified without republishing them.
- The declared token budget is enforced, not advised. In the default guaranteed mode the compiler fills to a margin below the stated figure and then removes the lowest-ranked passages until the rendered pack fits.
- Every response carries machine-readable non-claims: selection is extractive, evidence retention is best-effort, no claim verification is performed, completeness is not guaranteed, hallucination prevention is not guaranteed, and token counts are model-neutral estimates rather than provider billing counts.
- A reviewer holding the same inputs can recompute both hashes and confirm the pack is the one the evidence describes.

FAIL-CLOSED BEHAVIOUR

Each row below is a local contract test against the interceptor and the WSO2 Interceptor Service v1 policy implementation. None of it is a statement about your deployed environment.

- Missing interceptor configuration returns HTTP 503 and forwards nothing upstream.
- A missing or invalid interceptor credential returns HTTP 401, is compared in constant time, and is stripped from the request rather than forwarded.
- A decoded body above 512,000 bytes is refused with HTTP 413 rather than truncated.
- With the interceptor timing out or refusing the connection, the gateway policy returns an immediate error and does not call the model. This was verified against the WSO2 policy implementation itself, not a Maha reimplement of it.

- The response phase verifies an HMAC seal over the request-phase evidence. Missing or tampered evidence fails closed instead of returning a successful response.
- A request that does not carry the explicit Maha extension passes through untouched.

BUDGET, EXPANSION AND PAYMENT

- Below 1,024 estimated tokens the interceptor forwards your original context instead of a compiled pack, and above it compares the two and forwards the original whenever compilation would not be smaller. Enabling the policy therefore cannot increase the context sent to your provider.
- The decision is declared in the response headers, so a bypass is visible rather than silent.
- The published evaluation runner performs no automatic retries and requires an explicit spend ceiling before it may make a paid call.
- Context control performs no payment action. The compiler signs nothing, holds no key and authorises no transfer; payment, where it applies at all, is a separate module on a separate resource path and is not part of the interceptor flow.

KNOWN LIMITATIONS

- Maha's published evaluation corpus is synthetic. It establishes nothing about behaviour on your documents.
- Maha holds no security certification or regulatory attestation, and this document is not one.
- Maha is not a WSO2 partner and claims no WSO2 endorsement, certification, approval or customer validation. The integration is independent compatibility work.
- No saving, latency, availability or provider behaviour is guaranteed, and no result is promised before measurement.
- Selection is ranking and de-duplication. It does not verify claims and does not protect against prompt injection, data exfiltration or a hostile document.
- Yours to operate: the gateway-side credential mechanism, TLS termination, network egress, provider retention settings, and your own request logging. The published policy bundle is labelled evaluation-only because the interceptor-call credential is a gateway header rather than a reviewed secret reference.

VERIFY IT YOURSELF

```
npm test -- test/wso2-context-interceptor.test.ts
    fail-closed, credential stripping, bypass and seal behaviour
```

```
npm run validate:context-control-security-boundary
    every claim here maps to a committed source whose bytes still match
```

```
npm run reproduce:wso2-evaluation
    the frozen evaluation, as a dry run that contacts nothing
```

No credential is needed for any of the above, and none of them contacts a gateway, a model provider, or any Maha production system.