Authorizing Official one-pager — HR Benefits Assistant

System: HRBA · **Impact level:** Moderate · **Deployment:** agency-self-hosted

> **What this document is.** Evidence and derived documentation coverage, generated from the system profile and the control statements this repository gates on every build. It supports a risk-acceptance decision. **It is not an assessment result and asserts no control outcome.** No statement here means a control was tested and passed; that determination belongs to an assessor and to you.

---

1. Data flows

3 boundary crossing(s). **1 leave(s) the authorization boundary.**

	Path	Crosses	Data categories	Enforced by
`TB1`	Application to proxy	Internal — within the authorization boundary	employee-pii, benefits-enrollment, free-text-inquiry	mTLS; stream identity resolution
→ `TB2`	Proxy to external model endpoint	OUT of the authorization boundary	benefits-enrollment, free-text-inquiry	Deny-by-default endpoint allowlist with jurisdiction constraint; detection and enforcement applied before forwarding; TLS 1.3 re-origination
`TB3`	Proxy to evidence store	Internal — within the authorization boundary	decision-metadata	Hash-chained records; no prompt content by default

`TB2` is the crossing the security value of this change rests on. MATERIALLY CHANGED — this crossing previously had no enforcement point. It is the crossing on which the security value of this change rests.

2. Endpoints

Listeners added by this change:

Listener	Port	Protocol	Source	Authentication	Purpose
`lowrisk-proxy`	8443	TCP/TLS 1.3	hrba-app (in-boundary)	**mTLS**	Application-to-proxy AI request path
`lowrisk-proxy`	9090	TCP/HTTP	in-boundary monitoring only	**none**	Health and Prometheus metrics. No request content exposed.
`lowrisk-enrichment`	8081	TCP/HTTP	lowrisk-proxy (in-boundary)	**mTLS**	Sampled asynchronous analysis queue

> `lowrisk-proxy:9090` takes **no authentication**. Bound to the pod network only; not exposed at the boundary.

Changed: Application egress to the external model provider (TCP/443) no longer originates from hrba-app. It originates from lowrisk-proxy. The protocol and port are unchanged; the source principal is not.

3. Controls applied

49 control(s) are carried by at least one requirement the software implements today. Requirements marked *planned* are excluded.

- **49** have a written implementation statement — what the software does, what proves it, and what it does not reach.

0 of 180 requirements are *verified* (61 are implemented). In this repository *verified* means a named human signed off against evidence. Tests passing is not that, and the count is printed here rather than omitted because it is the first thing an assessor will ask.

The control-by-control table is **Annex A**, which is part of this package rather than an appendix. It is not reproduced here because a one-pager that runs to 49 table rows stops being read.

4. Residual risk, and its basis

4 risk(s) require explicit acceptance. Each states the basis for the claim, not only the claim.

RR-1

Detection is not complete. Published evaluations of mainstream PII and PHI detectors land broadly in the F1 0.6-0.9 range, with a characteristic 15-20 point drop when applied to a corpus other than the one they were tuned on. Sensitive data can pass undetected.

Mitigation. Customer-runnable evaluation harness measures recall on the agency's own representative data. Declared per-profile minimum recall and accepted false-negative rate. Sampled human review of allowed traffic.

Accepted by: System Owner — signature required — *not yet signed*

RR-2

Model-based detection (layer 3) is advisory and runs out-of-band. It cannot block. A finding it raises describes traffic that has already left the boundary.

Mitigation. Deterministic validators and the analyzer registry run inline and do block. Enrichment findings are recorded with provenance=enrichment and are never counted as enforcement actions.

Accepted by: System Owner — signature required — *not yet signed*

RR-3

The proxy is a new single point of failure on the AI egress path. Under the default fail-closed posture, proxy unavailability makes the AI feature unavailable.

Mitigation. Multi-replica deployment with pod anti-affinity; the data plane runs on its last valid configuration if the control plane is unavailable. Agencies may elect fail-open, which is recorded in configuration and surfaced in evidence as an accepted risk.

Accepted by: System Owner — signature required — *not yet signed*

RR-4

The FIPS build variant binds AWS-LC-FIPS 3.0.x, which is IN PROCESS with NIST CMVP, not yet validated. It is listed on the Modules In Process list.

Mitigation. Agencies requiring a citable certificate number deploy the 'fips-certified' variant, which binds AWS-LC-FIPS 2.0 (CMVP #4816).

Accepted by: ISSO — determination required before deployment — *not yet signed*

5. Failure behavior

What happens when a component is unavailable, and what the person using the system sees.

- **observeMode** — not visible to the user. No user-visible change. Traffic is recorded and analyzed; nothing is blocked or altered. This is the deployment posture for the first phase.
- **redactMode** — —. Detected values are replaced with placeholders before the prompt leaves the boundary, and restored in the response the user sees. In the normal case the user notices nothing. Where the model's answer depended on the redacted value, the answer may be less specific — for example, a question about a specific direct-deposit account may return general guidance.
- **blockMode** — visible to the user. The request does not reach the model. The user sees a reason code and generic text. The matched value is never echoed back (SI-11).
- **notificationPosture** — —.

  RR-3 is the availability risk: The proxy is a new single point of failure on the AI egress path. Under the default fail-closed posture, proxy unavailability makes the AI feature unavailable.

6. Coverage limitations

Generated from the `NOT COVERED` clause of each statement section 3 draws its coverage from — the same sentences, read twice. They cannot disagree, and this section grows whenever a coverage claim is added.

49 of 49 claimed controls carry a stated limitation. Each headline below is the first sentence of that control's clause; the full text is in **Annex B**. Nothing is summarised away — the annex is the same sentences at full length.

- `ac-3` — three privileged endpoints exist, because the administrative API is otherwise unbuilt.
- `ac-4` — CUI detection here (REQ-DET-5) recognizes the marking SYNTAX only — the `CUI//` banner/inline and `(CUI)` portion-mark forms — and is deliberately NOT a CUI determination: a determination designates information as controlled and is a government authority a syntactic match does not hold, so a bare unmarked `CUI` is refused, and a multi-word dissemination control is spanned only to its first space pending a registered marking dictionary.
- `ac-6` — grants are static configuration.
- `ac-6.9` — the privileged functions covered are the ones that exist: `signals:read`, `inventory:read` and `evidence:erase`. Policy writes, enforcement advancement and grant management have no endpoints, so there is nothing to log; they inherit this mechanism when they land, because `required_permission` denies by default and the recording is not a step those endpoints can omit.
- `au-10` — this is non-repudiation against an EXTERNAL forger, not against the deployment itself: the signing key lives in the proxy's process, so the log WRITER holds it and could re-sign a rewrite.
- `au-11` — the retention PERIOD is a customer determination recorded in their SSP. Windowed disposal of a log that HAS held a legal hold is refused pending a full-file hold replay (decision 0077, increment 2), and the raw/redacted tiers have no data to dispose until content capture exists (REQ-TEL-4).
- `au-12` — configuration-change events do not exist because configuration is a file read once at startup with no runtime mutation path; a change is a restart, and the restart discards the audit store, which is the AU-4 defect and not this one.
- `au-2` — requests that establish NO principal are counted in `lowrisk_admin_unauthenticated_total` and deliberately not chained (decision 0026), because an unauthenticated write path into an unbounded in-memory store is a denial-of-service surface and the record would name nobody; a credential-stuffing run against the admin listener is therefore visible as a count and not as an attributed trail, and alerting on that counter is a compensating control rather than monitoring hygiene.
- `au-3` — on DATA-PLANE records the "who" is still a stream identifier or a client-certificate subject rather than an authenticated principal; attribution to a person there requires principal resolution, which does not exist (decision 0013).
- `au-3.1` — the additional information does not include the payload or any excerpt of it.
- `au-4` — `queue_depth`, the gauge that would let an operator see that backlog before it becomes shedding, has no consumer.
- `au-5` — a genuine DEVICE error (a full disk, EIO, a revoked mount) is not exercised by a test; it shares the sticky-unhealthy path that the capacity refusal does prove, and inducing a real one portably needs a

filesystem fixture rather than an injected writer that would test the injection.

- `au-6` — **the review and analysis are still somebody else's.** This forwards records; it does not correlate them, alert on them, or define what a reviewer should look for.
- `au-6.1` — the integration is one-directional.
- `au-6.3` — correlation itself happens in the customer's SIEM, not here, and `reliably_ordered` currently has no caller because the console timeline that would consume it is unbuilt.
- `au-7.1` — the offering states dispositions and coverage; it does not assert that a control is validated, and the reconciliation proves the two responsibility sources agree, not that either is correct.
- `au-8` — discipline is operator-ATTESTED, not measured: the product records and holds the operator to naming the clock they declare, but does not itself query the host's NTP or PTP sync state, so a false attestation yields authoritative- looking timestamps the acknowledgement attributes to whoever signed it.
- `au-9` — a verification record proves that this deployment checked its own chain and says nothing about whether the deployment is honest.
- `ca-6` — **this product authorizes nothing.** CA-6 is an organizational control performed by a senior official; what is implemented here produces an INPUT to that decision.
- `cm-6` — identity settings this control is often expected to cover do not exist: there is no federated SSO, no group-to-role mapping from IdP claims, no SCIM, and no session policy (REQ-OPS-8a–8d, all planned).
- `cm-7` — the role and tenant logic is enforced at runtime for exactly three endpoints (`/signals`, `/whoami`, `POST /evidence/erase`); everywhere else in `lowrisk-authz` it remains a library this product ships and does not use, and each unwired entry point is named in compliance/disconnected-allow.yaml.
- `ia-2.1` — the privileged surface is three endpoints; the rest of the administrative API does not exist, so most of the model still has no caller (see compliance/disconnected-allow.yaml).
- `ia-2.2` — account provisioning, password policy, certificate issuance and factor enrolment are entirely the customer's identity infrastructure; this is authorization over an assertion someone else produced, not an authentication system.
- `ia-3` — two of REQ-CLS-1's four named criteria are wired (header, route); API-key-prefix and JWT/OAuth-claim classification are not, so a deployment relying on those cannot yet classify by them.
- `ia-5.2` — the distinction is acted on by exactly one action: `POST /evidence/erase` demands recency, so a card-authentication certificate is refused there outright and a PIV Authentication certificate is refused unless an air-gap waiver is configured.
- `ia-8` — authentication only; the same grant check governs what the principal may then do, and no grant means nobody.
- `mp-6` — there is no bulk or query-based erasure, no way to find the sequence number except by reading the chain, and no undo (correctly — the design is a tombstone); windowed disposal of a log that HAS held a legal hold is refused pending a full-file hold replay.
- `ps-3` — screening itself is performed by the organization employing the person and is a customer responsibility in the CRM. The review surface has no production caller; nobody can run it today.
- `ps-7` — this is a shared-responsibility statement, not a technical control.
- `pt-2` — determining the authority to process, the lawful basis, and the applicable retention period is the customer's determination in their System Security Plan and privacy documentation.
- `pt-3` — stating the processing purpose, obtaining any required consent, and publishing the applicable notice are customer obligations in the CRM. LowRisk.io detects markings and patterns; it does not determine CUI status, make export-control determinations, or de-identify health information, and any statement that it does is wrong.
- `ra-3` — **the product does not assess risk.** It transcribes risks a human authored and derives limitations from statements a human wrote.
- `sa-11` — the gate proves the record documents a pass, not that the pass was thorough or honest; reading that is review, and third-party penetration testing (CA-8) of the deployed boundary is a separate, unbuilt evaluation.
- `sa-8` — `derive_tenant` has NO PRODUCTION CALLER. The proxy has no authenticated principal to derive a tenant from, so cross-tenant assertion is not merely unenforced — it is not expressible.
- `sc-13` — the 3.0.x line is a Modules-In-Process entry, not a certificate.
- `sc-23` — this is DEFENSE IN DEPTH reading an UNVERIFIED claim: the proxy holds none of the issuer's signing keys and does not verify the token signature, so it catches naive / confused-deputy passthrough of a GENUINE mis-audienced token, NOT a forger who mints a token with a matching `aud`. Correctness of the token remains the MCP server's responsibility; this is not an authentication mechanism.

- **sc-4** — a session id is a caller-supplied CORRELATOR: isolation holds for DISTINCT ids, but two callers who present the SAME id share a map by definition, as with any session token.
- **sc-5** — **this is the only denial-of-service bound in the product, and it covers one operation.** There is no request rate limit on the data plane, no concurrency cap on proxied requests, no connection limit and no body-size limit outside the administrative erase endpoint's 16 KiB. A deployment relies on the surrounding infrastructure — ingress, service mesh, load balancer — for all of those, and this statement should not be read as a claim that the proxy protects itself from volumetric attack.
- **sc-7** — the proxy protects the boundary it is placed on and cannot see traffic routed around it.
- **sc-7.5** — the exception workflow's grant record is produced by the evidence layer; the interface through which a human would REQUEST and APPROVE an exception is part of the unbuilt admin API, so today an exception is expressed by editing configuration and restarting.
- **sc-8.1** — the default `fips` variant binds AWS-LC-FIPS 3.0.x, which is IN PROCESS, not certified.
- **si-10** — the token SIGNATURE is not verified (see the sc-23 statement); this is claim-level input validation, not authentication.
- **si-11** — the acknowledgement attributes the DEPLOYMENT's choice at load time in configuration; it is not re-stamped onto each individual disguised response, so an investigator reading one compatibility-mode record sees the discrepancy flag and resolves the accepter through the configuration, not the record.
- **si-12** — the retention PERIOD is a customer determination recorded in their SSP. A deployment that has not configured `evidence_log.retention` disposes nothing on a timer; the raw/redacted tiers have no data to dispose until content capture exists (REQ-TEL-4), and disposal does not yet reach a windowed file.
- **si-12.3** — the admin-configurable ceiling covers the audit-metadata tier only (the raw/redacted tiers await REQ-TEL-4).
- **si-4** — detection accuracy is UNMEASURED (REQ-ACC-7): no recall figure or false-positive rate is published because none has been measured, and the `accuracy` configuration block that would declare budgets is refused at startup precisely so a declared budget cannot be mistaken for a measured one.
- **si-4.2** — CUI detection in this real-time path is marking-syntax only and not a CUI determination; recovering a bare unmarked `CUI` and the full multi-word dissemination-control name are a Layer-2-context and marking-dictionary follow-up (decision 0064).
- **si-7** — the signing key lives in the proxy's own process, so it defends against an EXTERNAL tamperer who reaches the file at rest, NOT against the log WRITER, which holds the key and could re-sign a rewrite; a non-extractable HSM/KMS key is the increment that resists the writer.
- **sr-11** — a fingerprint suppresses one finding at one commit and cannot un-disclose the value, which stays in the history permanently; and D7 checks that a sentence was written, never that it is true.

Acceptance

Signing accepts the residual risks in §4 on behalf of the agency, on the basis of §§1–3 and §§5–6. It does not assert that any control was assessed.

Role	Name	Signature	Date
Authorizing Official			
System Owner			
ISSO			

Generated by `scripts/gen_ao_onepager.py` from `HRBA`'s system profile, `compliance/traceability.yaml` and `compliance/control-statements.yaml`. Regenerate rather than edit: a hand-edited copy stops matching the repository it claims to describe, and nothing will tell you.