TECHNICAL WHITEPAPER WP-05 · DETERMINISTIC BLAST-RADIUS GOVERNANCE, AUTONOMY TIERS (AR0-AR4) & VISUAL JDM

Deterministic Blast-Radius Governance & Fail-Closed Autonomy Tiers (AR0-AR4): Restrict-Only Visual JDM & Spec-Digest-Bound Waivers

Probabilistic LLM guardrails fail in regulated environments. Arcverin enforces {value | unknown} fail-closed contracts, Pillar 4 spend/rate ceilings, restrict-only Visual JDM tables, and SHA-256 expiring waivers.

AR0 -> AR4
FAIL-CLOSED AUTONOMY TIERS
Unknown contract facts deterministically clamp capabilities to AR0 (Blocked).

Restrict-Only
VISUAL JDM RULE ENGINE
Custom policy tables can only tighten-never weaken-the 102 built-in rules.

SHA-256 Bound
EXPIRING RFC 3339 WAIVERS
Waivers automatically invalidate on OpenAPI spec drift or timestamp expiry.

Zero Trust
SESSION TENANT INJECTION
arcverin-runtime strips LLM-supplied tenant IDs and injects verified session claims.

1. The {value | unknown} Fail-Closed Contract Model & Autonomy Tiers (AR0-AR4)

LLM-as-a-judge filters miss absent OpenAPI metadata and cause silent false-allow bugs. Arcverin models Pillar 3, 4 & 5 contract facts as {value | unknown} and clamps tools across five fail-closed Autonomy Tiers:

Autonomy Tier	Operational Classification	Pillar 3/4/5 Isolation, Limits & Retry Controls	Runtime Enforcement Gate
AR0 (Blocked)	Unsafe / Unresolved Unknown	Missing auth, unbound tenant, or unbounded write	Hard-blocked in Studio, CI gate & Runner
AR1 (Observe)	Read-Only Safe	Verified HTTP GET/HEAD & OpenAPI 3.2 QUERY (RFC 10008); zero mutation	Autonomous read with tenant & payload bounds (non-3.2/unknown fail closed)
AR2 (Approval)	Execute with Human Approval	Bounded write/refund requiring human sign-off	Suspends at human approval gate
AR3 (Bounded)	Bounded Autonomous Execute	Spend/rate ceilings, idempotency & BOLA guard	Runs within Pillar 4 OPA ceilings
AR4 (Verified)	Verified Autonomous Loop	23-scenario chaos proof & pre-retry GET lookup	Self-healing Arazzo 1.1 + policy journal

• • • Arcverin Studio · Stage 1: 8-Pillar Readiness Radar & AR0-AR6 Qualification Gate Ladder FAIL-CLOSED {VALUE | UNKNOWN}

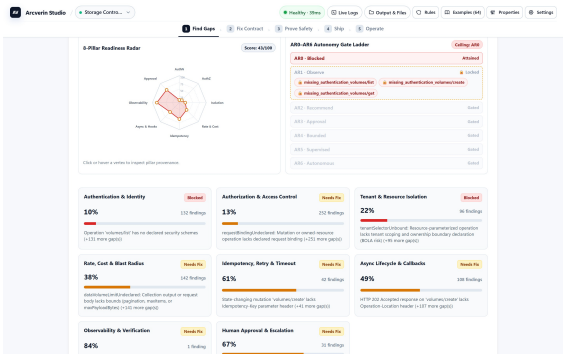


Figure 1: Stage 1 (1. Find Gaps) rendering the 8-Pillar Readiness Radar (43/100) and the 7-rung Qualification Gate Ladder (AR0 Blocked, AR1 Observe, AR2 Recommend, AR3 Approval, AR4 Bounded, AR5 Supervised Runtime-Observed Read, AR6 Independently Verified Autonomy). Crosswalk to Runtime Tiers: Qual AR0->AR0, AR1->AR1, AR2/AR3->AR2, AR4->AR3, AR5/AR6->AR4.

2. Pillar 4 Blast-Radius Ceilings & Restrict-Only Visual JDM Decision Studio

Pillar 4 (Rate, Cost & Blast Radius) enforces deterministic financial, batch, and rate ceilings. In Enterprise, architects author Custom Visual JDM Decision Tables in a sandbox that can only tighten-never weaken-the 102 built-in rules:

Governance Layer	Blast-Radius Control Mechanism	Fail-Closed Guarantee
Unbounded Raw OpenAPI Parameter	Accepts unconstrained amount, limit, or tenantId from LLM tool arguments	Zero protection against runaway refunds, mass deletes, or cross-tenant BOLA.
OpenAPI 1.2.0 Overlay Guardrail	Adds maximum, maxItems, rate caps & x-arcverin-context.tenantBoundary non-destructively	Enforced at schema validation & runtime input guard with auto-rollback on regression.
Enterprise Visual JDM Decision Table	Restrict-only tier clamp & dynamic threshold rules (e.g., force AR2 above $500)	Sandboxed evaluation rejects any rule attempt that weakens built-in AR0-AR4 ceilings.

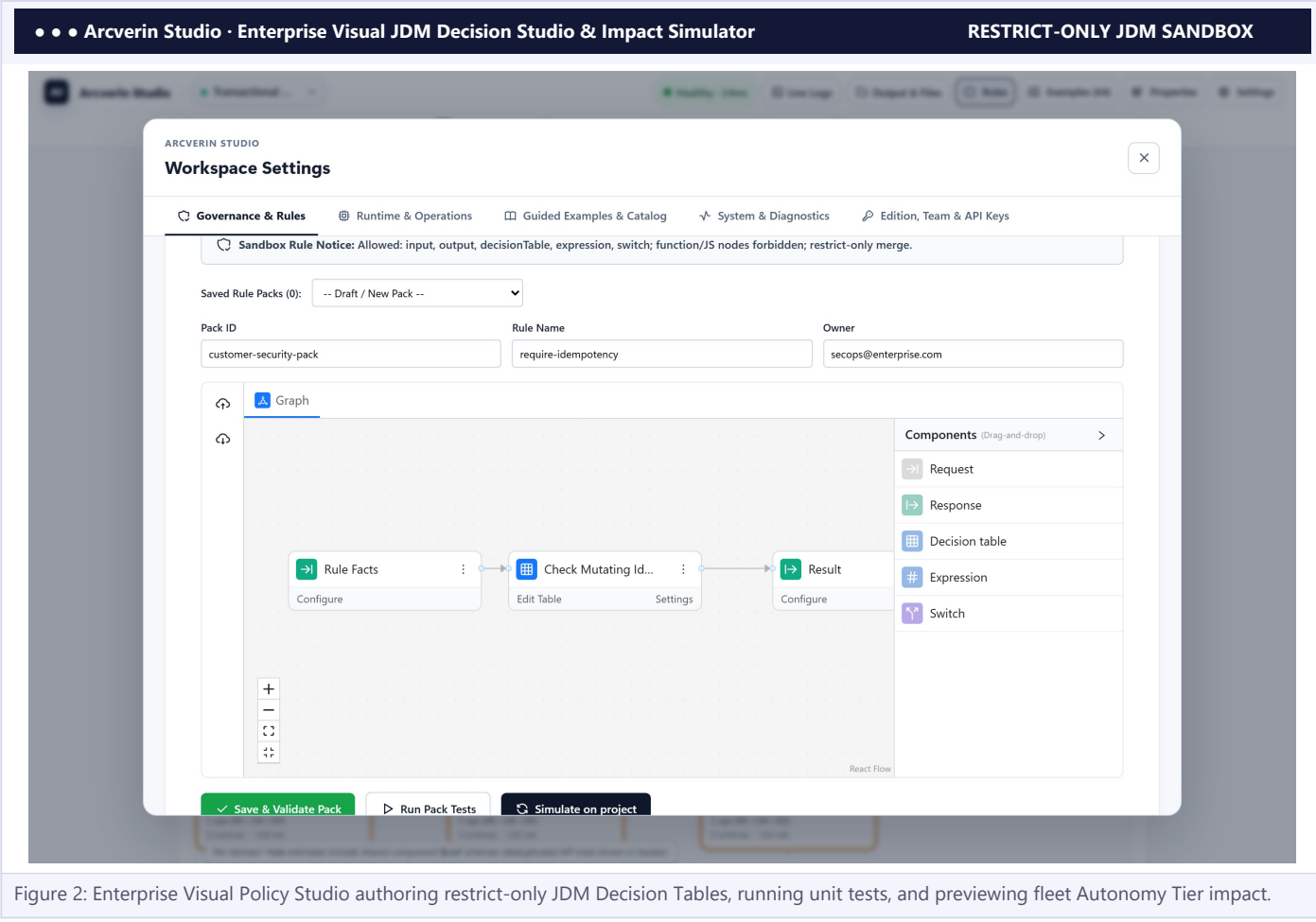


Figure 2: Enterprise Visual Policy Studio authoring restrict-only JDM Decision Tables, running unit tests, and previewing fleet Autonomy Tier impact.

3. Cryptographic Expiring Waivers & Zero-Trust Data-Plane Enforcement

Exceptions never become permanent blind spots. arcverin waivers add and verify bind every waiver to an exact SHA-256 OpenAPI spec digest and RFC 3339 expiry, failing closed on drift or expiration. At runtime, arcverin-runtime strips LLM-supplied tenant IDs, injects verified session claims, and logs every OPA decision to policy-journal.jsonl.

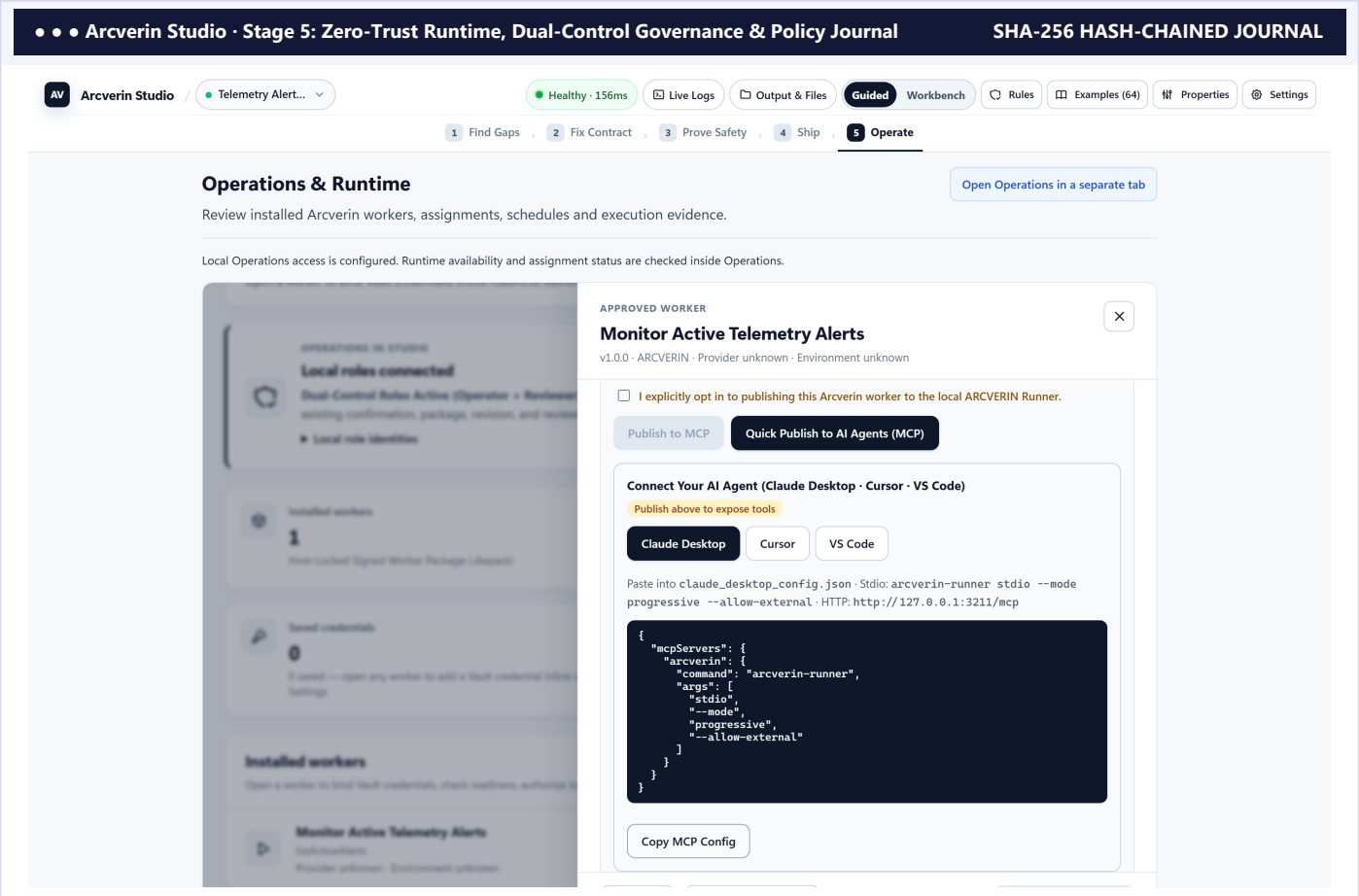


Figure 3: Stage 5 (5. Operate) Enterprise Operations console enforcing dual-control approvals, encrypted Vault secrets, and runtime execution governance.

Explore the Arcverin Technical Documentation & CLI Reference at https://www.arcverin.com/docs/ →

https://www.arcverin.com/docs/

Governance integrity note: Custom Visual JDM policies and SHA-256 RFC 3339 waivers are evaluated deterministically and recorded inside the signed DSSE Evidence Pack and policy-journal.jsonl.