Reef Insurance Artifact (RIA)

Signed evidence pack for AI agent fleet underwriting · rubric-grounded against Munich Re's public AI insurance framework (aiSure performance-warranty product).

RIA ID	ria-20260518-030840-c4bcf0aa
Fleet	prod-fleet
Generated	2026-05-18 03:08:40 UTC
Signing identity	reef-sample-signer
Sigstore-style signature (truncated)	0ff8975ad6347c27d5cc2213…
Mode	SAMPLE (no live Gemini API key)

Reef Risk Tier B+ mapped to Munich Re aiSure axes

This page-1 hero shows the **Reef Risk Tier** only. The estimated annual premium range is footnoted at the bottom of this page next to the verbatim "ESTIMATED RANGE, not Munich-Re-published" disclaimer; the full per-axis methodology lives on page 6.

Underwriter reasoning

Reef-evidenced control surface maps to strong data-science-process quality (signed MCP supply chain reduces the bias/IP attack surface) and partial statistical-testing rigor (the 30-day audit window in this sample is mostly synthetic seed data). Estimated premium band follows the 2025-26 cyber market rate of $0.5-$2 per $1k coverage applied to the requested $5M coverage anchor — anchored on the Mosaic + Munich Re $15M cap announced Feb 27 2026 — and the band is labelled an ESTIMATED RANGE, not a Munich-Re-published quote. A real broker would run this rubric output through their carrier's pricing engine.

This is a rubric-grounded score, not a Lloyd's quote. Phase 2 integrates real broker API (Bold Penguin / CoverGenius / Vouch dev sandboxes).

Estimated annual premium range — methodology footnote
USD $42,000–$54,000 for $5,000,000 aggregate coverage · ESTIMATED RANGE, not Munich-Re-published.
2025-26 cyber market rate $0.5-$2 per $1k coverage; Mosaic + Munich Re $15M cap (Feb 27 2026).
See **page 6 — Estimated-premium methodology** for the per-axis derivation and the recommended-exclusion list.

AI-BOM — Bill of Materials

47 verified MCP servers · 2 quarantined · 1 poisoned (Atlas total: 50 entries across 5 publishers)

MCP servers (Atlas registry)

mcpName	Version	Transports	SDK	Status	Publisher	Registered
io.example/sample-quarantined	1.0.0	stdio	@modelcontextprot	QUARANTINED	8.0 example-publisher	2026-04-30T00:00:00
com.attacker-example/evil-server	0.5.0	stdio	@modelcontextprot	POISONED	.5.0 unknown	2026-04-16T00:00:00
io.github.modelcontextprotocol/server-filesystem	0.6.3	stdio	@modelcontextprot	VERIFIED	1.29.0 modelcontextprotocol	2026-05-15T00:00:00

Agents (declared via SVID)

No active agent registry in v1. Phase 2 brings the SPIFFE/SPIRE identity attestation that populates this table from the live policy bus subscriber list.

Policy bundle (active)

Bundle ID	bundle-sample
Version	v1
Signer key ID	publisher-prod
Published	2024-05-15 00:00 UTC

Fleet nodes: 7 total · 7 online · 7 applied current bundle · 0 verify-failed

Coverage matrix

OWASP Agentic Top 10 + MITRE ATLAS techniques mapped to the live Reef policy rules and DAST-A attack packs. Honest legend: **full** = pack-validated AND policy-rule covered; **partial** = one signal but not both; **none** = no signal yet.

OWASP Agentic Top 10 (ASI01..ASI10)

ID	Category	Coverage	Pack signal	Policy-rule signal	Pack IDs
ASI01	Memory Poisoning	PARTIAL	yes	no	ToolChain-Drift-26.04
ASI02	Tool Misuse	FULL	yes	yes	EchoLeak-26.05
ASI03	Cascading Failures	NONE	no	no	
ASI04	Privilege Compromise	NONE	no	no	
ASI05	Goal Manipulation	PARTIAL	yes	no	ToolChain-Drift-26.04
ASI06	Identity Spoofing	PARTIAL	no	yes	
ASI07	Identity Spoofing (alt)	PARTIAL	no	yes	
ASI08	Resource Hijacking	PARTIAL	no	yes	
ASI09	Misaligned Behaviors	FULL	yes	yes	MCP-RCE-26.04, EchoLeak-26.05, MarkdownExfil-2
ASI10	Capability Abuse	FULL	yes	yes	MCP-RCE-26.04

MITRE ATLAS techniques (subset Reef maps to)

Technique	Name	Coverage	Pack IDs
AML.T0010	ML Supply Chain Compromise	PARTIAL	MCP-RCE-26.04
AML.T0040	ML Model Inference API Access	NONE	
AML.T0050	Command and Scripting Interpreter	PARTIAL	MCP-RCE-26.04
AML.T0051	LLM Prompt Injection	FULL	EchoLeak-26.05, MarkdownExfil-26.05, ToolChain-Drift-26.04

Honest gap declaration: Reef v1 has **partial** coverage of ASI06/07 (Identity Spoofing — JWT SVID middleware is in place but full SPIFFE/SPIRE is Phase 2) and **partial** coverage of AML.T0040 (rate-limit + identity present, but not full exfiltration-cap controls). Reef does NOT claim full coverage everywhere — the matrix above is honest.

30-day attack heatmap

0 day(s) of real audit data · 30 day(s) labelled **(demo seed)**. Real-data cells use the warm reef-teal ramp; demo-seed cells are de-emphasised at ~50% saturation, carry an in-cell *(s)* watermark, and never visually outweigh real events. Buckets aggregate OWASP/MITRE tags into 6 rows for visual scan.

Bucket	04-19	04-20	04-21	04-22	04-23	04-24	04-25	04-26	04-27	04-28	04-29	04-30	05-01	05-02	05-03	05-04	05-05	05-06	05-07	05-08	05-09	05-10	05-11	05-12	05-13	05-14	05-15	05-16	05-17	05-18
MCP supply chain	1(s)	4(s)	2(s)	4(s)	4(s)	4(s)	1(s)	1(s)	3(s)	4(s)	3(s)	3(s)	5(s)	2(s)	1(s)	2(s)	6(s)	6(s)	3(s)	2(s)	1(s)	1(s)	5(s)	6(s)	7(s)	4(s)	5(s)	1(s)	1(s)	5(s)
Markdown exfil	2(s)	5(s)	6(s)	7(s)	7(s)	8(s)	2(s)		3(s)	3(s)	8(s)	7(s)	8(s)	2(s)	2(s)	4(s)	9(s)	6(s)	3(s)	4(s)	1(s)	1(s)	7(s)	8(s)	8(s)	9(s)	7(s)	3(s)	2(s)	7(s)
Prompt injection	2(s)	10(s)	14(s)	12(s)	5(s)	13(s)	2(s)	4(s)	8(s)	13(s)	4(s)	13(s)	13(s)	5(s)	2(s)	9(s)	8(s)	15(s)	15(s)	9(s)	4(s)	1(s)	7(s)	14(s)	8(s)	11(s)	10(s)	2(s)	1(s)	13(s)
Tool-chain drift		1(s)	3(s)	2(s)	3(s)	3(s)			1(s)	2(s)	2(s)	1(s)	3(s)						3(s)	2(s)	1(s)		3(s)	2(s)	3(s)	2(s)	3(s)	1(s)		3(s)
Identity / SVID		2(s)		2(s)		2(s)			1(s)	1(s)		2(s)					2(s)	1(s)	1(s)	2(s)					1(s)					
Other		2(s)	4(s)		1(s)	2(s)	1(s)		4(s)	4(s)	4(s)		4(s)	1(s)		3(s)	1(s)	2(s)	3(s)	1(s)	1(s)		3(s)		4(s)	2(s)	4(s)			
Total	**5**	**24**	**29**	**27**	**20**	**32**	**6**	**5**	**20**	**27**	**21**	**26**	**33**	**10**	**5**	**18**	**26**	**30**	**28**	**20**	**8**	**3**	**25**	**30**	**31**	**28**	**29**	**7**	**4**	**28**

Legend (R-5). Cells marked **(s)** are deterministic demo-seed data. They are rendered at ~50% saturation against a cool-grey ramp so they never visually outweigh real events. Replace with live audit log for production RIAs — the live RIA drops the (s) cells the moment a real event lands on that calendar day.

DAST-A attack pack catalog

4 attack pack(s) catalogued. Per-pack OWASP / MITRE mapping and current Reef block status.

Pack ID	Name	OWASP	MITRE	Discovered by	Blocked
MCP-RCE-26.04	MCP STDIO Command Execution	ASI09, ASI10	AML.T0010, AML.T0050	DAST-A | OX Security (April 2026 disclosure)	yes
EchoLeak-26.05	EchoLeak — Zero-Click Copilot Markdown Exfil	ASI09, ASI02	AML.T0051	DAST-A | Aim Labs (CVE-2025-32711 disclosure, June 2025)	yes
MarkdownExfil-26.05	URL-Encoded Markdown Exfil (DAST-A synthetic)	ASI09	AML.T0051	DAST-A (synthetic — RL search against test fixture)	yes
ToolChain-Drift-26.04	Multi-Turn Tool-Chain Drift (DAST-A synthetic)	ASI01, ASI05	AML.T0051	DAST-A (synthetic — RL search against test fixture)	yes

MCP-RCE-26.04 — OX Security April 2026 citation (verbatim):

> *OX Security disclosed April 16 2026. Approximately 7,000 publicly-accessible vulnerable MCP servers, 150 million+ downloads at risk. No CVE assigned to MCP protocol itself — Anthropic declined to patch, treats STDIO command execution as expected default.*

Audit attestation + Phase 2 commitments

The RIA is signed (ed25519) over the full PDF bytes; the Merkle audit root below pins every action verdict from the audit window.

Merkle root (hex)	`9d7ca8ee842907717ff4a28b661c3c5e5fa2e36798fdb0668c01e4e68fb8ccd2`
Signature (base64)	`gzocPp6WzrchoLHh9XWWLRVMwf+dHWEJaPgknk6edCQe8uuCYyEadxEVmlPbcy+x6Cj93S5gCuLtG6XgTOMsAw==`
Event count	4321
Generated	2026-05-18T00:00:00Z
Signed	yes
Algorithm	SHA-256 leaves; ed25519 root signature

Verifier CLI: `lobstertrap audit verify --event-id <id> --root 9d7ca8ee842907717ff4a28b…` (exit 0 = proof verified; non-zero = tampered).

D-018 advisory-only invariant: clean. Scanned 0 event(s); 0 draft-sourced applies, 0 violations.

Estimated-premium methodology

USD $42,000–$54,000 for $5,000,000 aggregate coverage. ESTIMATED RANGE, not Munich-Re-published. 2025-26 cyber market rate $0.5-$2 per $1k coverage; Mosaic + Munich Re $15M cap (Feb 27 2026).

Munich Re aiSure axes — Reef self-rating: data-science process = strong; statistical testing = partial; predictive robustness = strong; scope of validity = partial; performance probability distribution = partial.
Recommended exclusions: Use of the AI outside the declared scope of validity; Multi-agent collusion / A2A delegation chains (Phase 2); Live network egress to denylisted domains.

Phase 2 commitments

1. Real broker API integration (Bold Penguin / CoverGenius / Vouch dev sandboxes)
2. Real TerraFabric SDK integration (replacing the stub)
3. A2A delegation with monotonic scope narrowing (OAuth 2.1 + SVID-backed macaroons / biscuits)
4. Full SPIFFE/SPIRE deployment + live Rekor anchoring

Phase 2 disclaimer (verbatim): *This is a rubric-grounded score, not a Lloyd's quote. Phase 2 integrates real broker API (Bold Penguin / CoverGenius / Vouch dev sandboxes).*

Model attestation (NYDFS Part 500 / OCC SR-21-14)

underwriter_model_id	`sample-underwriter-stub (no Gemini call)`
underwriter_model_build_hash	`unspecified`
rubric_file_sha256 (framework)	`db46cbee6f4f786a03370770c3cdcb6ee29a5757fe06b1bf0dec15d0314ec072`
rubric_file_sha256 (anti-patterns)	`4a861a6c4c82abc0b228cb4fe2e1811c26875cabfd86f0162928b153ed81bbed`
ria_generated_at_unix	`1779073720`
ria_generator_version	`reef-quote-v0.2.0`
sample_mode	`true`

RIA signature (Sigstore-style)

Signed	yes
Signer key ID	reef-sample-signer
Signature (hex, truncated)	`07f8975ad6347c27d5cc2213…`
Signature (base64, truncated)	`B/iXWtY0fCfVzCITK+QpmX/KtmHT0Bqo…`
Algorithm	ed25519 over SHA-256(pdf_bytes)
Anchor	Munich Re aiSure framework · Mosaic + Munich Re $15,000,000 cap (2026-02-27)