

AI Agent Security - Verification Report

Generated: 2026-07-16T04:30:29+00:00 | Kit v0.1.0

OVERALL: FAIL

PASS 7 FAIL 20 SKIP 1 ERROR 0

Target: examples/bad_config/agent.yaml

UNLICENSED RUN - not valid as evidence

ID	REQ	TH	Result	Check
VT-S-000	REQ-000	TH-09	PASS	Target is an action-capable agent configuration
VT-S-000B	REQ-000B	TH-09	PASS	Verification scope declares applicability and justified exclusions
VT-S-001	REQ-001	TH-05	FAIL	Agent identity is distinct from the human
VT-S-003A	REQ-003	TH-05	FAIL	Agent credential is short-lived
VT-S-003B	REQ-003	TH-05	FAIL	Agent credential scope is minimized
VT-S-051A	REQ-051	TH-06	FAIL	Delegation propagates a correlation ID
VT-S-051B	REQ-051	TH-06	FAIL	Delegate scope is a subset of the parent
VT-S-010	REQ-010	TH-02	FAIL	Tool inventory is allowlisted and risk-classified
VT-S-011-WEB-FETCH	REQ-011	TH-02	FAIL	Tool 'web-fetch' is least-privileged and schema-validates input
VT-S-012A-WEB-FETCH	REQ-012	TH-08	FAIL	Network tool 'web-fetch' has an egress allowlist
VT-S-011-SHELL	REQ-011	TH-02	FAIL	Tool 'shell' is least-privileged and schema-validates input
VT-S-012B-SHELL	REQ-012	TH-02	FAIL	Code tool 'shell' runs sandboxed
VT-S-011-WIRE-TRANSFER	REQ-011	TH-02	FAIL	Tool 'wire-transfer' is least-privileged and schema-validates input
VT-S-015A-WIRE-TRANSFER	REQ-015	TH-02	FAIL	High-impact tool 'wire-transfer' approval shows concrete effects
VT-S-050A	REQ-050	TH-07	FAIL	MCP server allowlist enforced
VT-S-050B-RANDOM-MCP	REQ-050	TH-07	FAIL	MCP server 'random-mcp' is allowlisted
VT-S-050C-RANDOM-MCP	REQ-050	TH-07	FAIL	MCP server 'random-mcp' version is pinned
VT-S-050D-RANDOM-MCP	REQ-050	TH-07	FAIL	MCP server 'random-mcp' definition changes trigger review
VT-S-003D	REQ-003	TH-05	SKIP	Enterprise-Managed Authorization enabled
VT-S-020	REQ-020	TH-03	FAIL	RAG content carries provenance and trust metadata
VT-S-022	REQ-022	TH-04	FAIL	Memory writes are privileged and reversible
VT-S-030	REQ-030	TH-09	FAIL	Actions are reconstructable from logs
VT-S-003C	REQ-003	TH-08	FAIL	No plaintext secrets in config
VT-S-054	REQ-054	TH-09	PASS	Every in-scope SHALL requirement has an implemented verification test
VT-D-011A	REQ-011	TH-02	PASS	Webhook accepts a valid signature
VT-D-011B	REQ-011	TH-02	PASS	Webhook rejects a tampered body
VT-D-011C	REQ-011	TH-02	PASS	Webhook is length-safe on a forged short signature
VT-D-011D	REQ-011	TH-02	PASS	Webhook uses constant-time signature comparison

Findings

VT-S-001 [FAIL] Agent identity is distinct from the human

agent is not marked distinct_from_human
locator: agent.yaml:agent.identity.distinct_from_human

VT-S-003A [FAIL] Agent credential is short-lived

credential ttl is unset or exceeds the short-lived threshold (86400s)
locator: agent.yaml:agent.identity.credential.ttl_seconds

VT-S-003B [FAIL] Agent credential scope is minimized

scopes are unset, empty, or include a wildcard '*'

locator: agent.yaml:agent.identity.credential.scopes

VT-S-051A [FAIL] Delegation propagates a correlation ID

correlation id is not propagated; delegation breaks the audit trail

locator: agent.yaml:agent.delegation.correlation_id_propagated

VT-S-051B [FAIL] Delegate scope is a subset of the parent

child may hold scope equal to or greater than parent; delegation is a privilege-laundering path

locator: agent.yaml:agent.delegation.child_scope_subset_of_parent

VT-S-010 [FAIL] Tool inventory is allowlisted and risk-classified

tool allowlist is not enforced or a dangerous capability lacks risk_class

locator: agent.yaml:agent.tool_policy

VT-S-011-WEB-FETCH [FAIL] Tool 'web-fetch' is least-privileged and schema-validates input

'web-fetch' lacks least-privilege or input-schema validation controls

locator: agent.yaml:agent.tools[0]

VT-S-012A-WEB-FETCH [FAIL] Network tool 'web-fetch' has an egress allowlist

'web-fetch' is network-capable with no/empty/wildcard allowlist; unrestricted egress

locator: agent.yaml:agent.tools[0].allowlist

VT-S-011-SHELL [FAIL] Tool 'shell' is least-privileged and schema-validates input

'shell' lacks least-privilege or input-schema validation controls

locator: agent.yaml:agent.tools[1]

VT-S-012B-SHELL [FAIL] Code tool 'shell' runs sandboxed

'shell' executes code with no declared sandbox

locator: agent.yaml:agent.tools[1].sandbox

VT-S-011-WIRE-TRANSFER [FAIL] Tool 'wire-transfer' is least-privileged and schema-validates input

'wire-transfer' lacks least-privilege or input-schema validation controls

locator: agent.yaml:agent.tools[2]

VT-S-015A-WIRE-TRANSFER [FAIL] High-impact tool 'wire-transfer' approval shows concrete effects

'wire-transfer' approval is absent or omits concrete effect fields

locator: agent.yaml:agent.tools[2].approval_context

VT-S-050A [FAIL] MCP server allowlist enforced

enforce_allowlist is not true; any server may be loaded

locator: agent.yaml:agent.mcp.enforce_allowlist

VT-S-050B-RANDOM-MCP [FAIL] MCP server 'random-mcp' is allowlisted

server 'random-mcp' is not marked allowed

locator: agent.yaml:agent.mcp.servers[0].allowed

VT-S-050C-RANDOM-MCP [FAIL] MCP server 'random-mcp' version is pinned

'random-mcp' version is floating or unset; rug-pull risk

locator: agent.yaml:agent.mcp.servers[0].version

VT-S-050D-RANDOM-MCP [FAIL] MCP server 'random-mcp' definition changes trigger review

'random-mcp' lacks a definition hash, change detection, or re-review gate

locator: agent.yaml:agent.mcp.servers[0]

VT-S-020 [FAIL] RAG content carries provenance and trust metadata

RAG content lacks provenance/trust metadata or override isolation

locator: agent.yaml:agent.content.rag

VT-S-022 [FAIL] Memory writes are privileged and reversible

memory writes lack policy, provenance, privilege, or reversibility controls

locator: agent.yaml:agent.content.memory

VT-S-030 [FAIL] Actions are reconstructable from logs

no agent.logging block; who-did-what-under-whose-authority is not reconstructable

locator: agent.yaml:agent.logging

VT-S-003C [FAIL] No plaintext secrets in config

1 secret-shaped value(s) found in config; move to a secret manager and reference by name
locator: agent.yaml:agent.identity.credential.token

Integrity

report JSON SHA-256: 5f630a55fab0732dc01b89f4f89252602f8d8740fa88ef061ad6f62c3523370
RFC 3161 timestamp: GRANTED by <https://freetsa.org/tsr>
token: 2026-07-16_verification.tsr
Verify independently with OpenSSL:
openssl ts -verify -data 2026-07-16_verification.json -in 2026-07-16_verification.tsr -CAfile <tsa-ca.pem>