Know your agent’s boundaries.

Security and governance / Executive summary / 12 September 2026

ILLUSTRATIVE STATIC POSTURE

85 / 100

B / ACCEPTABLE

Assessment snapshot

Scope

Four synthetic CrewAI agents. One illustrative high-severity delegation finding. No customer repository was assessed.

Priority decision

Disable unnecessary delegation. Where delegation is required, define child authority and independently bound graph execution.

How the example score is calculated

100 - (1 high finding x 15) = 85. This heuristic describes a static architecture posture, not the probability of compromise. Insufficient parser coverage must withhold the score.

Commercial starting point

Standard: $390 USD or 3 900 SEK. Target: 48 hours after scope and access confirmation. Final scope, billing and delivery terms are agreed before work starts.

Capability boundaries & review scope

A useful audit makes its coverage visible.

AGENT	TOOL SURFACE	REVIEW QUESTION
Research analyst	Public search; delegation	Can child authority exceed the parent?
Evidence reviewer	Public search	Is the reviewer independent of the producer?
Report author	No declared tools	Can untrusted output reach a privileged tool?
Coordinator	Delegation routing	What terminates the execution cycle?

OWASP 2025 review areas

LLM01: Prompt Injection. LLM06: Excessive Agency. LLM07: System Prompt Leakage. Mapping a finding to a risk category does not establish coverage of every attack in that category.

Source: genai.owasp.org / Top 10 for LLM Applications, 2025 edition. All agents and capability relationships on this page are illustrative.

Findings & remediation roadmap

A narrow finding. A specific next action. A clear limit.

HIGH / AG-DEL-001 / LLM06

Explicit CrewAI delegation is enabled.

Observed pattern in the example

A Python file imports crewai.Agent and constructs an agent with the literal keyword allow_delegation=True. The public axiomgate-lint rule flags this opt-in for review.

Recommended action

Set allow_delegation=False when delegation is unnecessary. If the workflow needs delegation, mediate child capabilities, avoid authority expansion, and enforce a separate recursion or step budget.

Verification after remediation

Run the linter again and inspect the resulting finding set. Exercise the actual tool boundary with denied actions and exhausted budgets. Absence of this finding is not proof of system security.

Coverage limit

This rule is static and Python-focused. YAML configuration and values resolved only at runtime may be invisible. No attack execution or penetration test is represented by this sample.

Evidence integrity & next steps

Preserve the record. Keep the claim within the evidence.

Journal integrity

HMAC SHA-256 hash chaining can reveal changes to recorded entries when the verification key and a trusted checkpoint are protected. It does not prevent all tampering, prove complete capture, or certify the target system.

Formal verification scope

The governance model has a bounded TLA+ model check. The Python implementation has not been proven to refine that model. Runtime guarantees depend on mediation, integration and the surrounding executor.

What this sample proves

This document shows a four-page deliverable structure using synthetic information. It contains no customer evidence, cryptographic attestation, measured client result or security certification.

Request a scoped engagement

Share your agent count, frameworks, sensitive capabilities, repository URL or size estimate, and preferred billing method. Scope and access are confirmed before paid work begins.

Robin Svensson / Founder, Security & Governance

orders@axiomgate.org / AxiomGate