PUBLIC METHODOLOGY REFERENCE

AI Agent Assessment Framework

A transparent structure for reviewing the architecture and controls of AI agents deployed on AWS.

ASSESSMENT OBJECTIVE

Translate architecture evidence into specific, practical control decisions.

The framework supports technical judgment. It does not certify security, establish compliance, or replace hands-on testing.

THE EIGHT REVIEW AREAS

- **01 Security boundaries**
- **02 Identity and permissions**
- **03 Agent governance**
- **04 Reliability and recovery**
- **05 Auditability**
- **06 Operational complexity**
- **07 Scalability and cost exposure**
- **08 Tool and multi-agent coordination**

WHAT WE DELIVER

An evidence-based assessment report with observations, validation boundaries, actionable recommendations, alternative implementation approaches, and a prioritized 30/60/90-day roadmap.

EVALUATION AREAS / 01-03

Safety boundaries and decisions

Key questions and the types of evidence that may support findings.

01 Security boundaries

Understand where user instructions, model outputs, tools, networks, and data cross trust boundaries.

Review questions: Can untrusted content influence tool invocation? Are secrets and sensitive data constrained? Which operations are consequential?

Illustrative inputs: Architecture diagrams, threat models, tool definitions, selected data-flow diagrams.

02 Identity and permissions

Understand effective privileges for each agent, tool, application identity, and human actor.

Review questions: Are permissions scoped to the required action? Are write and read identities separated? Where do authorization checks execute?

Illustrative inputs: Sanitized IAM policies, role assumptions, service authorization patterns, tool-to-API access map.

03 Agent governance

Understand how allowed actions, approval requirements, exceptions, and policy changes are controlled.

Review questions: Who approves irreversible actions? What does approval bind to? What fails closed? Who owns and changes the policy?

Illustrative inputs: Policy inventory, review procedures, approval records, decision flow, escalation design.

EVALUATION AREAS / 04-06

Operations and evidence

A practical view of how agents behave when systems, operators, or integrations are unavailable.

04 Reliability and recovery

Determine how failures, timeouts, retries, state changes, and interrupted work are handled.

Review questions: Are writes idempotent? Are uncertain outcomes reconciled? Do retries bypass approval or duplicate effects?

Illustrative inputs: Workflow diagrams, error contracts, retry configurations, incident playbooks, representative failure logs.

05 Auditability

Determine whether important actions and their control decisions can be reconstructed.

Review questions: Can a request be linked to policy decision, actor, tool input, execution, and outcome? Who can alter audit data?

Illustrative inputs: Redacted event schemas, trace IDs, CloudTrail and application logging design, retention/access rules.

06 Operational complexity

Understand who operates the architecture and how changes are safely introduced.

Review questions: Are responsibilities documented? Is drift detected? Can teams roll back policy and infrastructure changes?

Illustrative inputs: Deployment topology, runbooks, change approvals, ownership matrix, incident and support model.

EVALUATION AREAS / 07-08

Growth and coordination

How scale and multi-step agent work affect control effectiveness.

07

Scalability and cost exposure

Evaluate workload boundaries and the controls relevant to demand and resource use.

Review questions: Are budgets and quotas defined? Do concurrency and rate limits protect tools? Can backlogs affect human review?

Illustrative inputs: High-level volume estimates, quota settings, queue design, resource envelopes, cost monitoring approach.

08

Tool and multi-agent coordination

Understand how execution authority and state cross tool and agent boundaries.

Review questions: Are handoffs authenticated? Can one agent inherit another's permissions? Are delegated tasks scoped and traceable?

Illustrative inputs: Tool contracts, orchestration maps, delegation rules, event schema, representative handoff flow.

AWS CONTEXT

Depending on the customer architecture, the review may consider relevant use of Amazon Bedrock, Amazon EKS, AWS IAM, AWS CloudTrail, AWS KMS, and surrounding application services. Mentioning a service does not mean every engagement includes a configuration audit of that service.

POSSIBLE IMPLEMENTATION PATHS

Recommendations may compare AWS-native controls, custom services, and third-party platforms. Stacksona Gate may appear as an optional approach where it fits. The assessment is available without a Gate purchase.

METHODOLOGY / REPORTING

How findings become a decision

Clear evidence labels and practical next steps are part of the deliverable.

FINDING RECORD

01 Observation	What is present, absent, or uncertain in the supplied material.
02 Evidence	Document, excerpt, configuration, or discovery statement supporting the claim.
03 Confidence / limit	Observed, customer-reported, or unverified; access and coverage limitations.
04 Consequence	A specific failure or loss scenario linked to the architecture.
05 Recommendation	A control or architectural change tailored to the situation.
06 Alternatives	Other credible routes, their prerequisites, and their ownership costs.
07 Priority / sequence	Dependency-aware next steps rather than a generic checklist.

REVIEW PROCESS

- Confirm scope, objectives, permitted evidence, and data-handling requirements.
- Use discovery calls to close information gaps and distinguish assumptions from facts.
- Review selected evidence, record validation limits, and draft the assessment.
- Present report findings and feasible paths in a final advisory conversation.

STANDARD-SCOPE EXCLUSIONS

No penetration testing, formal compliance certification, comprehensive source-code security audit, implementation work, or ongoing infrastructure management. Depth and evidence access are agreed per engagement.