SAGAI

Strategy, Architecture, Governance and AI

ADLC: the Agent Development Lifecycle

Governance and engineering controls for building, approving, operating and retiring autonomous AI agents.

13	9+G2a	12	8	19
phases	gates	principles	control layers	documents

THE LIFECYCLE | 13 phases, 9 gates plus G2a, four stages

STAGE 1 | PHASES 1 TO 4 — Plan

Settle the case, scope, risk and design before anything is built.

1 G1	2 G2	3 G2a	4 G3
Discover & Assess	**Define Intent & Scope**	**Risk, Data & Threat**	**Architecture & Controls**
Business case	Agent specification	Risk, data, threat model	Architecture, permissions
Mandatory	Mandatory	Mandatory	Mandatory

STAGE 2 | PHASES 5 TO 7 — Build & Evaluate

Generate, test and fix at speed; G4 escalates only on exceptions.

5	6	7 G4
Spec-Driven Build	**Generate, Test, Fix**	**Autonomous Eval Gate**
Versioned build spec	Candidate code, test logs	Eval report, security scan
Optional or on the loop	None, sandboxed	Automated, escalates

STAGE 3 — Govern & Release

Human sign-off, then release approval.

8 G5	9 G6
Human Checkpoint	**Release Readiness**
Approval, evidence pack	Rollback plan, config snapshot
Mandatory	Mandatory

STAGE 4 | PHASES 10 TO 13 — Operate & Improve

Run under watch, learn, and loop back to Plan.

10	11 G7	12 G8	13 G9
Deploy & Operate	**Feedback & Improve**	**Reassess**	**Retire**
Telemetry, drift flags	Change proposal	Reassessment report	Retirement record
Optional or on the loop	Mandatory	Mandatory	Mandatory

Governance shell. Humans hold every high-blast-radius decision, so risk, data and threats are settled before design.

Inner loop 6 ⇄ 7. Generate, test, fix inside a sandbox. No gates except G4's six exceptions.

Outer loop 10 → 11. Deploy, monitor, improve. A change proposal returns to Phase 2 or 5; a material change re-enters at G7.

Human involvement: Mandatory · Automated, escalates · Optional or on the loop · None, sandboxed

G5 sign-off scales with risk tier.

WHAT NEVER MOVES | six of twelve principles

- P1 **Humans stay accountable**
- P2 **Least privilege, per action**
- P3 **Every action traceable**
- P4 **Evaluation is permanent**
- P10 **Agent risk ≠ action risk**
- P9 **Gaps get logged, not dropped**

Also in force: Principle 8 (redact real names, employers, identifying role details) and Principle 12 (controls sit outside the agent).

RISK IS RATED PER ACTION

A0	**Read information**	Autonomous
A1	**Generate an artefact**	Automated evaluation
A2	**Modify non-production**	Automated gates
A3	**Modify shared or staging**	Approval or control
A4	**Modify production**	Human approval
A5	**High-impact, irreversible**	Human decision-maker

The stricter control always wins.

FIVE AUTONOMY LEVELS

- 0 **Advisory:** Info only. Reviews everything.
- 1 **Assisted:** Acts, needs confirmation each time.
- 2 **Supervised:** Executes predefined actions, spot-checked.
- 3 **Controlled Autonomous:** Independent within strict bounds.
- 4 **Highly Autonomous:** Coordinates multiple activities.

A default posture, capped per action.

EIGHT CONTROL LAYERS

1. **Constitution (AGENTS.md)** Static rules, loaded every session
2. **Knowledge modules** On-demand, per-domain context
3. **Hard gates** Deterministic, non-LLM checks
4. **Escalation interface** Routes specific failures to a human
5. **Eval telemetry** Continuous, not pass or fail once
6. **Memory / state** What it learned from prior runs
7. **Tool permissions** Least privilege, rated per action
8. **Runtime control plane** Limits, kill switch, external

LAYER 8: ENFORCED OUTSIDE THE AGENT

Agent

Policy enforcement point

Permission and action-risk check

Budget, rate, time check

Tool

Kill-switch flow: Detect → Contain → Disable → Preserve evidence → Investigate → Remediate → Re-evaluate → Re-authorise

G4: AUTOMATED, SIX FORCED ESCALATIONS

1. Security scan finding at or above threshold
2. Eval regression beyond tolerance
3. Tool, destination or data source outside the action catalogue
4. New or changed model or provider
5. Prompt-injection or goal-hijack pattern found
6. Repeated failure beyond the retry ceiling

GOVERNANCE OVERLAY, ALWAYS ON

- Agent Registry
- Risk Classification
- Info Class. & Privacy
- Threat Model
- Identity Model
- Supply Chain
- Legal & Regulatory
- Change Management
- Incident Management
- Versioning & Provenance
- Accepted Risk Log

Data & Privacy Assessment: Phase 3. The register checks it holds.

NINETEEN DOCUMENTS, FIVE GROUPS

FOUNDATION (2)
Framework Summary · Agent Spec Template

PER-AGENT BUILD (5)
AGENTS.md · Knowledge Module · Tool & Permission Manifest · Hard Gate Requirements · Runtime Control Plane

RISK & ASSURANCE (5)
Threat Model · Data & Privacy Assessment · Supply-Chain Assessment · Legal & Regulatory Assessment · Evaluation Pack

OPERATE & RESPOND (4)
Escalation Config · Incident Runbook · Change Mgmt / G7 Procedure · Human Checkpoint Checklist

CROSS-AGENT (3)
Agent Registry · Agent Identity Model · Info Classification Register

Sagai | ADLC Framework | CC BY 4.0, (c) 2026 Aquasila Pty Ltd