AI Agent Governance Checklist

A pragmatic checklist for responsible AI, security, and operational control.

Foundations

- Named owners for policy, security, and operational accountability.
- Clear data classification and retention rules (PII, PHI, confidential).
- Documented purpose: what the agent can and cannot do.

Access & Tooling

- Least-privilege tool permissions (allowlist per agent / per environment).
- Approval step for high-impact actions (payments, deletions, external comms).
- Secrets stored securely (no hardcoding), rotated on schedule.

Logging & Audit

- Structured logs for prompts, tool calls, and outcomes (redact sensitive data).
- Audit trails retained for required period; searchable for investigations.
- Incident playbook: who responds, escalation path, and rollback steps.

Evaluation & Monitoring

- Pre-launch evaluation suite: safety, correctness, and policy compliance.
- Continuous monitoring: drift, anomalies, safety flags, and cost spikes.
- Post-incident review process and update cadence for tests/guardrails.