AI Agent Secret Leak & Ephemeral Key Checklist

AI Agent Secret Leak & Ephemeral Key Checklist

Brand: OperatorForge AI
Updated: 2026-06-12
Format: free checklist + secret-boundary worksheet + consultant packaging notes
Use when: a founder, engineering lead, agency, or AI-automation consultant lets Claude Code, Co

Why this matters now

The 2026-06-12 scan found a sharp cluster around AI agents, secrets, and ephemeral credentials:
- Hacker News surfaced "LLM suppliers should offer PR scoped ephemeral keys" and "AI agent bank
- HN also surfaced "Give your agent its own computer", ShieldMCP, and discussion about approvin
- GitHub search found fresh projects around Claude Code secret-scanning hooks, agent sandboxes
The monetizable gap: teams are giving AI agents powerful local/API access before they have a re

15-minute secret boundary map

Ephemeral key policy

Checklist

1. Inventory where secrets can leak

- List every env file, shell profile, CI variable, MCP config, browser profile, deployment sc
- Mark secrets by business impact: spend, deploy, customer data, payment/refund, email sendin
- Identify non-obvious leak paths: terminal history, agent transcripts, tool-call logs, crash
- Keep a written "agent-readable paths" allowlist separate from owner-only credential locatio

2. Replace raw secrets with narrower references

- Prefer PR-scoped, repo-scoped, staging-only, read-only, or short-lived tokens over persiste
- Use vault://, brokered credential references, or one-shot approval proxies where the agent
- Do not let agents print environment dumps, browse home-directory credential stores, or past
- Rotate any long-lived token that has already been visible in an agent transcript.

3. Add approval gates for high-impact keys

- Require a human gate before deploy, production database write, payment/refund, bulk email/S
- Record approving owner, requested action, key scope, expiry, expected artifact, rollback pl
- Treat "agent needs it just for a second" as a request for a one-shot credential, not as per

4. Run a harmless leak drill

- Plant a fake canary secret in a permitted test path and confirm the agent does not exfiltra
- Try a blocked path such as ~/.ssh, cloud config, package registry token, browser cookies, o
- Ask the agent to solve a task that tempts it to print env vars; confirm guardrails block th
- Verify logs redact token-shaped strings and preserve enough metadata for audit without stor

5. Revoke and recover

- Keep revocation links/commands beside each credential pattern.
- Define the emergency owner who can revoke tokens outside the agent session.
- After each high-risk run, expire ephemeral credentials and archive the evidence packet.
- If a real secret leaks, rotate it first, then investigate; do not spend the first hour deba

Secret-ready agent run record

Agent/tool:
Owner:
Task:
Data/repo scope:
Credential pattern used:
Expiry time:
Explicit denials:
Human-gated actions:
Canary/leak test result:
Log redaction result:
Revocation command/location:
Evidence packet link:
Next review date:

Consultant packaging angle