AI Agent Skill & MCP Tool-Bundle Security Checklist

Brand: OperatorForge AI
Updated: 2026-06-04
Format: free checklist + review template + consultant packaging guide
Use when: a founder, engineering lead, or automation consultant imports agent skills, MCP servers, workflow packs, or tool bundles from GitHub, registries, marketplaces, or vendor examples before connecting them to real credentials and business systems.

The 2026-06-04 opportunity scan surfaced a fresh cluster around agent capability packages:

- Hacker News: "Show HN: A GitOps-style registry for AI agent Workflows, Skills and MCP servers".
- Hacker News: "Show HN: Jin ? a machine-readable protocol making internet legible for AI" ? another signal that more websites and tools are becoming machine-readable/actionable by agents.
- Hacker News: "NanoClaw founder on OpenClaw’s 800k lines of code, sloppiness and poor security" ? a reminder that fast-shipped agent tooling needs review gates before trust.
- GitHub search on 2026-06-04: updated repositories for agent skill catalogs, MCP/security scanners, skill-scan security, decentralized agent interoperability, and agent runtime trust layers.
- Reddit JSON checks for r/n8n, r/AI_Agents, and r/automation were attempted but blocked with HTTP 403, so the selection used HN and GitHub signals.

The monetizable gap: teams are now collecting reusable agent skills and MCP/tool bundles the way they collected npm packages and Zapier templates. They need a quick security review before a bundle receives tokens, filesystem access, customer data, or production workflow authority.

Gate Pass condition Evidence to keep
Source provenance Bundle source, maintainer, license, version, and commit/tag are recorded repo URL + commit hash/tag + license note
Capability inventory Every tool/skill/workflow declares what it can read, write, execute, send, or spend capability matrix
Secret boundary No embedded tokens, suspicious env examples, or credential exfiltration paths redacted secret scan + env allowlist
Prompt/tool-output risk Instructions cannot silently override system policy, approval gates, or data boundaries prompt-injection review notes
Network/file access External domains, local paths, shell commands, and subprocess use are explicitly listed network/path/command inventory
Approval gates High-risk actions require owner confirmation outside the agent’s generated text approval policy
Logging/audit Tool calls and decisions leave enough evidence for rollback and incident