



AI Plug-In Security Best Practices – vCISO Template

Objective: Safeguard organizational assets by managing risks associated with AI plug-ins (e.g., Claude Skills, GPT Tools, other automation plug-ins).

1. Plug-In Permissions

- Apply **least privilege principle**: grant only necessary access.
- Avoid “single-consent” approvals for critical systems or data.
- Review and revoke unnecessary permissions regularly.

2. Code Review & Vetting

- Mandatory security review before deployment.
- Scan for **hidden functions**, external downloads, or execution commands.
- Use automated static/dynamic analysis where feasible.

3. Approval Workflow

- Centralized approval by a **limited, trained group**.
- Maintain **audit logs** of plug-in approvals and updates.
- Require **multi-person approval** for high-risk plug-ins.

4. Isolation & Segmentation

- Run plug-ins in **sandboxed environments**.
- Restrict **network connectivity** to only essential endpoints.
- Limit file system access to **non-critical directories**.

5. Monitoring & Logging





- Real-time monitoring of file, network, and command activity.
- Alerts for unusual behaviors or script execution.
- Maintain logs for **incident response and audits**.

6. Employee Training

- Educate staff on AI plug-in risks and safe usage.
- Run simulations to test awareness of malicious plug-ins.
- Reinforce reporting channels for suspicious activity.

7. Incident Response

- Include AI plug-in threats in **incident response plans**.
- Have procedures for **quick isolation or removal** of malicious plug-ins.
- Coordinate with **IT and legal teams** during ransomware or malware events.

8. Vendor & Platform Security

- Use only **verified or trusted sources**.
- Keep AI platforms **patched and updated**.
- Review vendor security documentation and transparency reports.

9. Risk Assessment

- Include AI plug-ins in **periodic cybersecurity risk assessments**.
- Conduct **threat modeling** for AI attack vectors.
- Align with regulatory requirements (ISO 27001, NIST CSF).

Quick Governance Tip for vCISOs:

Treat AI plug-ins like any other enterprise software. Combine technical controls with policy, training, and monitoring. The goal is **safe adoption, not blind trust**.

