Keep agents inside a controlled boundary

Many agent risks do not start with the model. They start with where the agent runs, what it can reach, and how quickly access can be revoked when something goes wrong. If runtime is scattered across unmanaged devices, broad network paths, or loosely governed credentials, blast radius expands fast.

This checklist helps teams evaluate whether an agent workflow is contained well enough for real production use, especially where sensitive systems, regulated records, or high-value operations are involved.

Use this checklist when:

- Deciding whether an agent can run outside a pilot environment
- Reviewing runtime architecture and network boundaries
- Testing whether revocation, inventory, and egress controls are strong enough

Checklist

- ☐ **Centralized, governed runtime:** Does the agent run in a controlled environment rather than on employee laptops or unmanaged VMs?
- ☐ **No unapproved runtimes:** Are there technical controls preventing "agent in the wild" paths from executing the same workflows?
- ☐ **Workflow-scoped network access:** Can you allow only required applications and destinations, and restrict general internet access where appropriate?
- ☐ **Segmentation by risk:** Can you isolate high-risk workflows into tighter network and policy boundaries with stronger evidence retention?
- ☐ **Credential discipline:** Are credentials scoped to the minimum necessary, rotated, and protected from extraction or reuse outside policy?
- ☐ **Least privilege by default:** Is access granted per workflow and time-bounded, not persistent broad entitlements?
- ☐ **Fast revocation:** Can you revoke agent access instantly and terminate active sessions, with proof revocation occurred?
- ☐ **Asset inventory:** Can you enumerate all agents, their owners, permissions, and the systems they can access?
- ☐ **Change management:** Are agent configuration and permission changes controlled, reviewed, and logged?
- ☐ **Egress controls:** Can you prevent data from being sent to unauthorized external destinations, including "helper" services?