Kill switch and rollback for a security agent

AGENT INCIDENT PLAN for <agent name>
Version <date>, owner <role>, next drill <date>

1. Switch off now if any of these is true
[ ] A change failed its independent check and the agent proceeded
[ ] The audit log has a gap the agent cannot explain
[ ] A change touched a system outside the approved tool register
[ ] The agent's action count in 10 minutes exceeds <limit>
[ ] The model supplier announces an incident or a change we did not schedule

2. Who decides
[ ] Decision: <role>, deputy <role>. Anyone on call may stop first and ask afterward.

3. First 15 minutes
[ ] Disable the agent's credentials. Record the time.
[ ] Freeze the log: copy it to a store the agent cannot write to.
[ ] List every change from the register since the last good check.

4. Undo
[ ] Reverse changes in the order of the rollback register. Mark each as restored or compensated.
[ ] Name who reopens any findings the agent closed (a closed finding is a claim until the system confirms it).

5. Tell
[ ] Asset owners for each changed asset, within <hours>.
[ ] Escalate to <risk authority> when the criteria in section 1 are met.

6. Restart conditions
[ ] Cause written down, a test added to the regression set, sign-off by <role>.

Agentic AI in Cybersecurity, Kill switch and rollback for a security agent, edition 19 Nov 2026, https://agenticaiincybersecurity.com/evaluation-checklist/kill-switch-and-rollback/. License CC BY 4.0.