CYBERINSIGHTS | PRACTICAL SECURITY

Govern the action, not just the chatbot: a practical agentic AI control plan

A practical evidence aid. Not certification or an automated assessment.

1 REGISTER

Define business purpose, owners, permitted data and prohibited actions. Separate discovery from permission to operate.

2 MAP

Inventory model, retrieval, memory, identities, tool servers, destinations, queues and human review points.

3 LIMIT

Use scoped credentials and tool-side authorization. Treat retrieved material and model output as untrusted.

4 APPROVE

Show target, data, action and consequences. Bind authorization to exact arguments and reject changed or expired approvals.

5 TEST

Use synthetic data to verify allowed and denied actions, tenant separation, retries, cancellation and safe handling of misleading content.

6 STOP

Exercise execution pause, credential revocation, queued work control and in-flight reconciliation.

7 OPERATE

Monitor outcomes, minimize stored sensitive content and re-review changes to models, tools, identities and data sources.

Evidence: scope / owner / expected result / observation / gaps / reviewer / recheck

Sources: OWASP LLM01:2025: Prompt injection; OWASP LLM06:2025: Excessive agency; NIST AI Risk Management Framework; NIST AI 600-1: Generative AI profile. Full links and limitations are in the companion guide. Examples are hypothetical; no customer environment was tested.