AI agent assurance

Test tool privileges, human approval and the stop path together.

Printable workpaper. The detailed guide and all scenario branches are in the Markdown file.

Enterprise support, document processing and action-taking agents. Use synthetic data and inert tools in tests.

Process and expected evidence

1. Define action boundaries

Tool–permission–data map

2. Assess impact

Impact record and human decision point

3. Test trust boundaries

Test input, expected and observed result

4. Exercise approval and stopping

Approval-bound action and stop record

5. Reassess after changes

Version-bound evaluation and monitoring

Work record

Scope / date:
Business / technical owner:
Decision and rationale:
Evidence reference / retest:

Example and closure

Fictional example; not evidence of actual work.

Agent: AI-001

Tool / action: CRM write

Data: Synthetic customer

Permission owner: Support process owner

Approval condition: Destination + fields approved

Negative test: Destination change after approval

Observation: Action must be rejected

Action: Add to version tests

Closure checklist

[ ] Tool and data paths mapped.

[ ] Authorization enforced by the application.

[ ] Negative approval and stop tests exist.

[ ] Version changes trigger reassessment.

Official sources

NIST — AI Agent Standards Initiative

https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative

ISO/IEC 42001 — AI management system overview

https://www.iso.org/standard/42001

ISO/IEC 42005:2025 — AI impact assessment overview

https://committee.iso.org/standard/42005?browse=tc