

Ourbit Vendor-Neutral AI - Day 8: Tools and AI Agents

Day 8 - Tools and AI Agents

Programme: Vendor-Neutral AI - 15-Day Certification Programme

Purpose: Understand the difference between generating advice and allowing an AI-enabled system to take actions through tools.

Learning outcomes

- Explain a plan > act > observe loop.
- Distinguish model reasoning/output from tool execution.
- Apply least privilege to tool access.
- Define approval gates, stop conditions and logs.
- Recognize when autonomous action is inappropriate.

1. From answers to actions

A language model that drafts an email is producing content. A system that can call an email tool and send the message can change the outside world. This difference is fundamental. Once tools are available, permissions, validation and accountability become part of the AI design.

2. Plan > act > observe

A tool-enabled workflow may identify a next step, call an allowed tool, inspect the result and decide what to do next. The loop can be useful for research, data processing or operational tasks. It also creates opportunities for compounding errors if the system continues acting on a mistaken assumption.

3. Least privilege

Give the system only the capabilities necessary for its task. If it needs to read a calendar, it does not automatically need permission to delete events. If it may draft an email, it does not automatically need permission to send one. Separate read, create, modify, delete and approve permissions where the environment allows it.

4. Human gates and stop conditions

Consequential actions should have proportionate controls. A human approval gate can sit before payment, publication, deletion, account changes or external commitments. Stop conditions should end or escalate a run when required information is missing, a tool returns an unexpected result, permissions fail, risk rises or repeated attempts do not resolve the problem.

5. Logs create accountability

Record what task was requested, which tools were invoked, what material result came back, what consequential action was proposed or performed, and who approved it where approval was required. Logs should be useful for diagnosis without unnecessarily storing sensitive data.

Worked example - research and email agent

An agent may search approved sources, compile findings and draft a client email. It may not send the email. The final draft and source list are presented to a staff member, who verifies important claims and explicitly approves sending. This preserves automation benefits while keeping an external commitment under accountable control.

Guided practice

Design permissions for an assistant that manages meeting preparation. List what it may read, create and modify. Identify at least two actions it must not perform without human approval and define two stop conditions.

Knowledge check

- Why is tool use materially different from text generation?
- What happens in a plan-act-observe loop?
- What does least privilege mean?
- Where should human gates be placed?
- Why are logs important?

Applied evidence task

An AI assistant can read and edit customer records. Define the minimum permissions, validation rules, approval gates, logging and stop conditions required before allowing it to perform any write action. Explain which risks your controls address.

Day 8 takeaway

An AI agent should not be measured only by what it can do. Good design also defines what it cannot do, when it must stop, what requires approval and how its actions remain auditable.

Assessment boundary

This learning evidence belongs only to vendor-neutral-ai-15-day.

Learning comes before credentials. Completing this reading alone does not issue a certificate. Programme assessment and practical requirements apply.