

Ourbit Vendor-Neutral AI - Day 11: Responsible AI, Privacy and Security

Day 11 - Responsible AI, Privacy and Security

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

Purpose: Make responsible choices about data, permissions, consequences and accountability before and during AI use.

Learning outcomes

- Apply data minimization to AI tasks.
- Distinguish access permission from permission to disclose or process data elsewhere.
- Apply least privilege to connected tools.
- Identify consequential decisions needing human accountability.
- Design safer alternatives when information is sensitive.

1. Responsibility starts before the prompt

Before sharing information with any AI system, ask whether the information is needed, whether you are permitted to use it for this purpose, where it may be processed, and what could happen if it is exposed or misused. Technical access to data does not automatically create authority to disclose it.

2. Data minimization

Use the minimum information necessary to complete the task. Remove identifiers and unrelated details where possible. Prefer summaries, redacted extracts, synthetic examples or controlled internal systems when they satisfy the objective with less exposure.

3. Permissions and least privilege

Connected AI systems should receive only the permissions required. Read access is different from write access; drafting is different from sending; proposing a change is different from executing it. Narrow permissions reduce the impact of mistakes and misuse.

4. Human accountability

AI can support decisions, but consequential judgments involving rights, safety, employment, finance, legal obligations or other significant effects require proportionate human control. The accountable person should understand the evidence and limitations rather than merely rubber-stamping an AI recommendation.

5. Security is part of workflow design

Consider authentication, authorization, sensitive outputs, prompt or document injection, unsafe tool actions, logging and retention. Security controls should protect the entire workflow, not only the model endpoint.

Worked example - confidential case summary

A staff member wants AI to summarize a customer dispute. Instead of copying the entire account record into an unknown service, the workflow removes identifiers and unrelated financial data, uses only the facts necessary for summarization, confirms the approved processing environment and keeps the final decision with authorized staff.

Guided practice

Review a hypothetical personnel case file. Identify what an AI summarization task genuinely needs, what should be redacted, what should never be sent to an unapproved service, and who should approve any consequential next action.

Knowledge check

- What does data minimization mean?
- Why is access not the same as disclosure permission?
- What is least privilege?
- When should human accountability increase?
- Name two security considerations beyond the model itself.

Applied evidence task

A team proposes uploading a spreadsheet containing names, contact details, payment history and support notes to a public AI tool for trend analysis. Redesign the task to reduce exposure while preserving analytical value. State the approvals and controls you would require.

Day 11 takeaway

Responsible AI is operational: minimize data, control permissions, protect the workflow, understand consequences and keep accountable humans responsible for consequential decisions.

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.