

Ourbit Vendor-Neutral AI - Day 10: Evaluating AI Output

Day 10 - Evaluating AI Output

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

Purpose: Judge AI output by evidence and fitness for purpose rather than fluency or confidence.

Learning outcomes

- Evaluate accuracy, relevance, completeness and evidence.
- Separate facts, inference, recommendations and uncertainty.
- Recognize hallucination and unsupported specificity.
- Check for material bias or missing perspectives.
- Scale verification to the consequence of error.

1. Evaluation begins with the task

An output cannot be called good in the abstract. A creative brainstorm and a compliance briefing have different standards. Start with the task's success criteria, evidence requirements, audience and consequences, then evaluate the response against them.

2. A practical evaluation frame

Ask: Is it accurate? Is it supported where support is required? Is it relevant to the question? Is anything important missing? Are uncertainty and assumptions visible? Could framing or data create a material bias? Is the output safe and appropriate for its intended use?

3. Unsupported specificity

Precise numbers, dates, quotations, citations and named claims deserve attention because specificity can create false confidence. Treat unsupported precision as a verification signal, not as proof of authority.

4. Consequence-based verification

Low-stakes ideation may need only a quick reasonableness check. High-impact decisions require stronger evidence, independent verification and often qualified human review. Verification effort should rise with the harm that an error could cause.

Worked example - supplier recommendation

An AI recommends Supplier A because it is “30% more reliable.” Before acting, ask where the figure came from, what reliability means, what period was measured, whether comparable suppliers were evaluated and whether important criteria such as cost, security or service were omitted. A confident recommendation is not yet a defensible decision.

Guided practice

Take a short AI-generated recommendation. Mark each material statement as verified fact, unverified claim, inference, recommendation or uncertainty. Then write the minimum verification required before using it.

Knowledge check

- Why is fluency not a quality measure by itself?
- What is unsupported specificity?
- How do task criteria shape evaluation?
- Why should verification scale with consequence?
- How can missing information bias a recommendation?

Applied evidence task

An AI answer is mostly correct but contains one unsupported claim that could materially change a business decision. Decide whether the output is usable, define the verification step, and explain how the final record should distinguish the verified material from the unresolved claim.

Day 10 takeaway

Evaluate AI output as evidence-bearing work: test it against the task, verify important claims, expose uncertainty and increase scrutiny as consequences rise.

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.