

Ourbit Master AI - Day 10: Evaluating AI Output

Purpose: Apply accuracy, bias, hallucination-risk and fit checks to AI output.

Source/build status: Recovered Module 6 is the source.

Learning outcomes

- Evaluate AI output for accuracy, bias, hallucination risk and task fit.
- Separate verified claims from unsupported claims, inference and opinion.
- Scale verification effort according to the consequence of being wrong.
- Revise or qualify claims that cannot be adequately supported.

Core lesson - Evaluating AI Output

Evaluate AI output with four practical checks: accuracy-are factual claims supported; bias-is framing unjustifiably skewed; hallucination risk-are sources, quotations, statistics or details unsupported; and fit-does the output satisfy the task? Verification effort should rise with the cost of being wrong. Confident tone is not an accuracy signal.

Worked example

A generated market summary contains a precise growth percentage and names a report. Confirm both. If the source cannot be verified, remove or qualify the claim rather than preserving it because it sounds convincing.

Teaching / video direction

Use an Accuracy / Bias / Hallucination Risk / Fit review card on one polished answer. Separate supported claims from unsupported ones. Close by widening the lens to privacy, permissions and responsible data judgment-Day 11.

Hands-on practical exercise

- Take an AI-generated answer containing at least three factual claims.
- Mark each claim as verified, unverified, inference or opinion.
- Inspect the framing for important omissions or unjustified bias.
- Identify quotations, sources, statistics or precise details that require confirmation.
- Compare the answer with the original task and constraints for fit.
- Revise the output, removing or qualifying claims that remain unsupported.

Knowledge check

- Name the four evaluation checks used in this lesson.

- Why is confident tone not an accuracy signal?
- What should happen to a precise statistic that cannot be verified?
- What does the fit check evaluate?
- How should consequence affect verification depth?
- Why should supported facts be distinguished from inference?

Assessment and remediation contract

The scored daily quiz uses 10 questions; 7/10 (70%) is a pass. Missed concepts map to exact lesson anchors for targeted remediation, including on a passing attempt. Attempts 1-9 below threshold return to remediation and a fresh eligible question set; an unsuccessful tenth attempt restarts the day in a new learning cycle while preserving history. The 50-question validated bank is a learner-release gate and will be scaled only after the Day 1 runtime passes acceptance.

Production / QA notes

- Use the worked scenario as the primary visual sequence; avoid generic AI-brain imagery.
- Keep terminology vendor-neutral and define unavoidable technical terms in context.
- Do not equate fluent output with verified truth or AI assistance with autonomous authority.
- Before learner release, fact-check the final lesson and validate every scored item against explicit teaching.

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