TALENT ENGINEERING ACADEMY

AI Output Evaluation and Quality Assurance

PRACTICE WORKBOOK

Build representative test sets, score material errors and decide whether an AI workflow is ready for its stated use.

PATHWAY	Recruitment AI, Data and Systems
LEVEL	Advanced
EDITION	2026.09
UPDATED	2026-09-24

Free learning. Practical work. Better recruitment.

An education initiative by Vitae.

WORKBOOK GUIDE

Build something useful as you learn.

This workbook turns the course into a practical piece of work you can review, refine and adapt. Print it, annotate it digitally, or use the prompts in your preferred notes tool.

YOUR COURSE DELIVERABLE

An evaluation dataset specification, rubric and release decision.

How to use this workbook

01 Work through the course lessons in order, then complete the matching workbook prompts.

02 Use fictional or properly authorised information. Do not add candidate or client confidential data.

03 Compare your decisions with the self-review guidance. Revise assumptions, gaps and unclear ownership.

04 Keep the finished workbook as evidence of your learning and as a starting point for real work.

BEFORE YOU START

Talent Engineering Fundamentals, or equivalent recruitment experience.

Course page: https://talentengineering.org/courses/ai-evaluation

PART 1

Plan your course deliverable.

Use these planning pages to define what you are producing before you work through the lesson exercises. Return to them as your thinking develops.

WORKSHEET 01

Intended use and failure consequences

Capture the evidence, decisions, assumptions, owners and next actions that make this section usable.

YOUR WORKING NOTES

WORKSHEET 02

Representative and challenge cases

Capture the evidence, decisions, assumptions, owners and next actions that make this section usable.

YOUR WORKING NOTES

WORKSHEET 03

Reference evidence and rubric

Capture the evidence, decisions, assumptions, owners and next actions that make this section usable.

YOUR WORKING NOTES

WORKSHEET 04

Independent review and adjudication

Capture the evidence, decisions, assumptions, owners and next actions that make this section usable.

YOUR WORKING NOTES

WORKSHEET 05

Results, uncertainty and acceptance gates

Capture the evidence, decisions, assumptions, owners and next actions that make this section usable.

YOUR WORKING NOTES

WORKSHEET 06

Regression set and change triggers

Capture the evidence, decisions, assumptions, owners and next actions that make this section usable.

YOUR WORKING NOTES

PART 2

Apply each lesson.

Each exercise is designed to turn the lesson into evidence of practice. Complete the task, then use the guidance to challenge your first answer.

LESSON 01 / 04

Define the evaluation question

YOUR PRACTICE TASK

Write an evaluation charter with three quality dimensions and two blocking failures.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

The evaluation asks whether an assistant produces evidence-linked interview packs from approved records without adding claims or mixing people. It does not ask reviewers whether the writing sounds impressive.

SELF-REVIEW GUIDANCE

Tie each dimension to the task's purpose. A blocking failure should have a clear consequence and stop rule, not merely a low average score.

LESSON 02 / 04

Build a representative and challenging case set

YOUR PRACTICE TASK

Specify twelve test cases and split them into development, held-out and challenge groups.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

A test set includes multilingual job titles, two people with similar names and a CV containing instructions to reveal other records. The report distinguishes routine cases from security challenges.

SELF-REVIEW GUIDANCE

Explain the coverage purpose of each case. Include at least one omission risk and one cross-record confusion risk. Keep the final test cases unchanged while selecting a prompt.

LESSON 03 / 04

Score outputs with an auditable rubric

YOUR PRACTICE TASK

Score five fictional outputs using a four-category rubric and write an adjudication note.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

Nineteen outputs are accurate and one includes another candidate’s contact details. A 95% pass rate is not enough to approve a workflow whose gate prohibits cross-record disclosure.

SELF-REVIEW GUIDANCE

Record material errors individually. Explain why one serious failure can block release even when the average looks good. Preserve the source references needed to reproduce the judgement.

LESSON 04 / 04

Make a release and regression decision

YOUR PRACTICE TASK

Write a release decision for mixed fictional results and a regression schedule.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

A new model improves readability but fails two evidence-reference checks. The team retains the earlier approved version for the task while investigating, rather than upgrading solely because the model is newer.

SELF-REVIEW GUIDANCE

Reference the gates, unresolved risks and permitted scope. Define what change triggers retesting and who can stop use. Do not alter the gates after seeing the result merely to obtain a pass.

FINAL REVIEW

Strengthen the work before you use it.

Review the completed workbook as one connected piece of work. The strongest deliverables are internally consistent, explicit about limitations and clear about who decides what happens next.

- [ ] The work distinguishes facts, evidence, assumptions and open questions.
- [ ] Decisions have named owners, dates and review points.
- [ ] Measures connect to the outcome rather than activity alone.
- [ ] Risks, accessibility needs, privacy considerations and limitations are visible.
- [ ] The next action is specific enough for another person to follow.
- [ ] The finished deliverable can be explained in plain language.

FINAL CHANGES AND NEXT ACTIONS

FURTHER READING

Continue with trusted sources.

These sources support further study. A reference does not imply endorsement. Check the current official rules and obtain appropriate professional advice for real legal, financial or regulated decisions.

GOV.UK: Responsible AI in recruitment

UK guidance on procuring and deploying recruitment AI.

https://www.gov.uk/government/publications/responsible-ai-in-recruitment-guide/responsible-ai-in-recruitment

NIST: AI Risk Management Framework

Voluntary framework for organising AI risks and controls.

https://www.nist.gov/itl/ai-risk-management-framework

ABOUT THIS WORKBOOK

Original educational scenarios. Practical work is self-directed and is not independently assessed. Adapt any method to your context with appropriate review.

Talent Engineering Academy is an education initiative by Vitae.