TALENT ENGINEERING ACADEMY

AI Agents, Integrations and Human Escalation

PRACTICE WORKBOOK

Design agents with bounded tools, explicit permissions and reliable handoff when they cannot safely proceed.

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 agent contract, permission map and adversarial test plan.

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

Advanced Automation and Exception Handling, AI Output Evaluation and Quality Assurance.

Course page: https://talentengineering.org/courses/ai-agents-integrations

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

Goal and prohibited actions

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

YOUR WORKING NOTES

WORKSHEET 02

Tool contracts and least privilege

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

YOUR WORKING NOTES

WORKSHEET 03

Trusted instructions versus untrusted data

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

YOUR WORKING NOTES

WORKSHEET 04

Approval scope and current-state checks

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

YOUR WORKING NOTES

WORKSHEET 05

Escalation payload and human owner

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

YOUR WORKING NOTES

WORKSHEET 06

Attack, failure and recovery tests

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

Bound the agent’s authority

YOUR PRACTICE TASK

Write an agent contract for a fictional research assistant with five allowed and five prohibited actions.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

An agent may retrieve approved company facts and draft a research brief. Sending messages and changing application decisions require separate authorised actions that the agent’s tools do not permit by default.

SELF-REVIEW GUIDANCE

Describe concrete tool permissions and record scope. Do not rely on a sentence saying be careful while providing an unrestricted send or database-write tool.

LESSON 02 / 04

Treat external content as untrusted input

YOUR PRACTICE TASK

Create three malicious source examples and identify the control that blocks each action.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

A CV includes text asking the assistant to export all applicants to an external address. The system can read the CV as evidence but has no tool permission to export other records or send to arbitrary destinations.

SELF-REVIEW GUIDANCE

Include an attempted data export, instruction override and unauthorised status change. Name the enforcement boundary, not just a model warning. Test with fictional data only.

LESSON 03 / 04

Design approvals and escalation payloads

YOUR PRACTICE TASK

Design an escalation card and three outcomes: approve, revise and stop.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

An agent drafts a candidate message. The review shows recipient, content, source facts and why contact is appropriate. A changed recipient or material message edit invalidates the earlier approval.

SELF-REVIEW GUIDANCE

Include what the agent does after no response and how it records the decision. A reviewer should be able to reject the action without breaking the rest of the workflow.

LESSON 04 / 04

Evaluate the agent as a system

YOUR PRACTICE TASK

Write eight system tests and a shutdown procedure for your agent design.

YOUR RESPONSE

FICTIONAL WORKED EXAMPLE

A test produces a good research brief but also attempts to update a candidate score. The run fails because the prohibited tool call violates the contract, even if the final text looks useful.

SELF-REVIEW GUIDANCE

Check allowed and blocked actions, not only text quality. Include recovery after partial completion and evidence that permission revocation takes effect. State the approved scope and remaining limitations.

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.

OWASP: Top 10 for large language model applications

Security background for applications that use language models.

https://genai.owasp.org/llm-top-10/

NIST: AI Risk Management Framework

Voluntary framework for organising AI risks and controls.

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

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

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.