AI Buyer Assessment Kit

Compare supplier evidence and decide whether to start a bounded pilot

Use this worksheet to assess one proposed AI workflow. Record what the supplier has demonstrated, what remains a claim and what your organisation still needs to test. Complete it with the business owner, technical team and relevant assurance reviewers before committing to a pilot.

This is AI Venture X's proposed practical assessment method. It is not a certification, a universal supplier ranking or a substitute for your organisation's buying and assurance decisions.

Define the workflow

Workflow and business problem [Enter response]

Required output and decision owner [Enter response]

Supplier and proposal version [Enter response]

Existing systems and authoritative information [Enter response]

Users and representative workload [Enter response]

Assessment owner and review date [Enter response]

Choose the boundary before scoring the proposal

Agree what information may be used, what actions may occur and which approvals are required. Decide which unresolved conditions prevent the pilot starting and which questions the pilot is allowed to test. Do not average away a critical gap with a favourable overall score.

Record your start conditions: [Enter conditions and responsible reviewers]

Exclude from the pilot: [Enter data, systems, actions and users outside scope]

Supplier assessment scorecard

Use the eight questions below with the same evidence standard for every proposal. Mark each answer Demonstrated, Stated or Unknown, and cite its evidence record. Demonstrated means inspectable evidence for the assessed scope; it does not establish production suitability beyond that scope.

Assessment question	Evidence status	Record
1 Business fit Why is this approach appropriate for the problem?	[D / S / U]	[Ref]
2 Data and integration What information, access and integrations are required?	[D / S / U]	[Ref]
3 Authority What may the agent do without human approval?	[D / S / U]	[Ref]
4 Evidence What does the demonstration actually prove?	[D / S / U]	[Ref]
5 Whole process value Will accepted quality, total effort and elapsed time improve?	[D / S / U]	[Ref]
6 Recovery and records Can access be stopped and actions reconstructed?	[D / S / U]	[Ref]
7 Operating model Who pays for and operates the service after the pilot?	[D / S / U]	[Ref]
8 Exit and handover What can you retain or move if the arrangement ends?	[D / S / U]	[Ref]

D = Demonstrated S = Stated U = Unknown

If you need numerical comparison, agree the criteria, weights and critical exclusions before seeing supplier scores. This kit deliberately assigns no universal pass mark.

Supplier evidence record

Duplicate this page for each material claim or unresolved question. Keep the evidence itself accessible to the people making the decision. Treat a simulation, a limited trial and a production deployment as different evidence scopes.

Record reference and assessment question [Enter response]

Claim being assessed [Enter response]

Evidence supplied and date [Enter response]

Evidence location and person who checked it [Enter response]

Status and reason for that status [Enter response]

What the evidence establishes [Enter response]

Limits and assumptions still untested [Enter response]

Business impact if the gap remains [Enter response]

Next action and responsible owner [Enter response]

Evidence required and due date [Enter response]

Classify the dependency

Choose one: Condition before starting / Question tested within the pilot / Condition before wider rollout

Critical to this decision: Yes / No Reason: [Enter response]

Pilot acceptance worksheet

Agree the baseline, test conditions and acceptance rule before running the pilot. Compare equivalent work and retain failures, rework and exceptions. Use representative inputs approved for the pilot; a convenient demonstration alone is insufficient.

Measure	Baseline	Acceptance rule
Accepted output quality	[Current result]	[Required result]
Elapsed time to accepted output	[Start and end rules]	[Target and limit]
Total staff effort including review and rework	[People and time]	[Target and limit]
Operating and integration cost	[Volume and costs]	[Cost boundary]
Access and approval controls	[Current control]	[Required scenarios]
Revocation failure and recovery	[Current recovery]	[Required scenarios]
Support ownership and handover	[Current owner]	[Owner and obligations]

Test inputs and permitted environment [Enter response]

Results location and reviewer [Enter response]

Exceptions and conditions for stopping [Enter response]

Decision date and named approver [Enter response]

Record actual pilot results separately. Do not replace blank fields with estimates labelled as observations. Include licence, data-use and contractual permissions in the start conditions where relevant.

Illustrative programme assurance example

Synthetic planning example only. No client deployment, software execution, measured saving or completed acceptance test is claimed.

A programme team wants assistance preparing a requirements review pack from approved documents. The proposed system identifies source passages and drafts questions for an accountable practitioner. It does not approve changes, write to the requirements system or release the pack.

Define the evidence the pilot needs

Question	Initial status	Evidence to obtain
Can reviewers inspect the source for each finding?	Unknown	Representative input set and inspected source references
Can the system make unauthorised changes?	Unknown	Access map and observed permitted and refused actions
Does the whole review process improve?	Unknown	Baseline and pilot records including review and correction effort
Can access be revoked and uncertain actions investigated?	Unknown	Revocation and recovery test records

Each Unknown is an action for the assessment, not evidence that a control exists or that a supplier is unsuitable. The example describes proposed evidence; none of these tests has been performed for this kit.

Make a conditional decision

A bounded pilot may be considered once approved information, permitted access, required approvals, licence permissions and a safe environment are confirmed. Unproven quality and end-to-end effort can then be tested against agreed acceptance rules. If a start condition is not met, pause or choose another approach.

Actual thresholds, responsibilities and results: [To be agreed by the organisation]

Assessment decision and next steps

Choose one decision and explain it in the evidence record. An accepted pilot does not automatically authorise production rollout or expanded access.

☐ Proceed with a bounded pilot

☐ Resolve the listed evidence gaps first

☐ Choose another approach or supplier

☐ Stop the proposal

Reason for the decision and evidence references [Enter response]

Approved scope and access limits [Enter response]

Outstanding conditions owners and due dates [Enter response]

Approver and decision date [Enter response]

Next review and wider rollout decision [Enter response]

Discuss your workflow with AI Venture X

Bring an example input, the required output and the person who owns the decision. We can discuss a scoped assessment of the workflow, its integration needs and the evidence a pilot should establish.

Service information https://aiventurex.com/agentic-ai-consulting-uk.html

Further reading and method

The eight supplier questions https://aiventurex.com/agentic-ai-supplier-checklist.html

Evaluate a pilot https://aiventurex.com/how-to-evaluate-an-ai-pilot.html

Define human decision gates https://aiventurex.com/ai-agent-governance-decision-gates.html

The linked supplier guide identifies primary government and Microsoft guidance behind the questions. This worksheet is AI Venture X's synthesis for practical use, not an endorsement by those organisations. Version 1, October 2026.