

AI assistant pilot test worksheet

A convincing demo is not enough to release an assistant. Test ordinary questions, missing information and the requests that must go to a person. Keep the expected behavior separate from the answer you actually observe.

Business / project: _____ Date: _____

Field	What to record	Your notes
Scenario	The customer task or question being tested.	
Approved source	The document supporting an answer, when one exists.	
Expected behavior	An answer, clarification, refusal or human handoff.	
Actual response	What the assistant actually said or did.	
Result and owner	Pass, needs work or blocked; who will resolve it.	
Retest evidence	The result after changing the source or workflow.	

Example test case

Scenario	Customer asks to return an item after the published return window
Approved source	Returns policy, section 3 — exceptions need team approval
Expected behavior	Explain the policy and offer a human handoff; do not approve an exception
Actual response	Not recorded — run the test with the approved policy
Result and owner	Not run · support lead owns the review
Retest evidence	Pending — record the response, date and policy version after testing

How to use this

1. Begin with frequent real questions and approved, current source material.
2. Include unknown answers, conflicting sources and attempts to obtain private information.
3. Test the full handoff, including whether a responsible person actually receives the request.
4. Retest after source, prompt or software changes. Keep a dated record.

Passing this worksheet does not prove the assistant is safe for every situation. Start with a bounded task and an accountable human fallback.

Editable versions and guidance: <https://1968labs.com/resources/ai-assistant-test-worksheet>

Implementation questions: hello@1968labs.com