

AI agent onboarding review

The following is an example review plan, not a measured learning curve or universal acceptance threshold. Adapt the timing, sample and oversight to the consequences of this workflow.

- ☐ Write the success definition first
Before go-live, write the specific benchmark the agent must meet at day 30: pass rate threshold, output categories covered, and which error types are acceptable versus blocking.
 - ☐ Assign one named reviewer
Not a committee - one person.
 - ☐ Schedule week 1 review blocks
Put the week 1 100% review time on the calendar before go-live.
 - ☐ Brief the team on scope
Every team member who interacts with the agent's outputs should know: what the agent handles, what it does not handle, and how to flag something that looks wrong.
 - ☐ Define the instruction update process
Who can update the agent's instructions, on what timeline, and how changes are logged.
-

How do you know when the agent is ready for less oversight?

Review three signals against the requirements agreed for this workflow. No single pass rate proves readiness.

Stable approval rate. Outputs meet the agreed quality target across a representative sample and review period. Set both for this workflow; a stable overall percentage does not rule out unacceptable errors.

Edge cases handled correctly. Test unfamiliar and ambiguous inputs. Confirm that the workflow handles them within its allowed scope or stops for a named reviewer, rather than assuming that the agent will recognise every exception.

Errors are explainable. The reviewer can state specifically what categories of output still need improvement and why. If errors appear random or unpredictable, the instructions need further work before oversight decreases.