

Production readiness checklist

A pilot is about to go live and you want to know if it should.

What has to be true before an AI pilot is allowed to go live. Agree the list before the pilot starts, so the go-live decision is a check, not a debate. Every item needs an owner and evidence. "We think so" is not evidence.

1 • Ownership

- ☐ A named person owns the system in production, and has agreed to.
- ☐ A named person owns the decisions the system supports. The system drafts, checks or summarizes. It does not decide.
- ☐ The people who use it were involved in the pilot and know what changes for them.

2 • Value and baseline

- ☐ The baseline was measured before the pilot started.
- ☐ The same measure was taken during the pilot, the same way.
- ☐ The result meets the exit criteria agreed at the start. If the criteria changed, the change was written down and agreed.

3 • Evaluation

- ☐ There is an evaluation set: real examples of the work, with the right answers, including the hard and unusual cases.
- ☐ The system passes it at the agreed threshold, and the results are saved.
- ☐ The evaluation runs again after every change to prompts, models, data or code, before the change ships.
- ☐ Someone who does the work reviewed a sample of real outputs, not only test results.

4 • Human review and escalation

- ☐ It is written down which outputs a person must check before they count.
- ☐ The reviewer can see what the system used to produce the output.
- ☐ There is a clear route when the system can't or shouldn't answer: it says so and hands off.

5 · Failure handling

- ☐ Known failure types are listed, with how each is caught.
- ☐ Wrong or harmful outputs can be flagged by users, and flags reach the owner.
- ☐ If the system is down, the work can continue the old way. People know how.

6 · Data, privacy and security

- ☐ The data it touches is listed, and its use is allowed under the organization's policies and any contracts.
- ☐ Access is limited to the people and systems that need it.
- ☐ The terms of any AI provider used were reviewed for data retention and training use.
- ☐ Logs keep what is needed to investigate a problem, and no more.

7 · Cost

- ☐ The cost per use is known, and a monthly limit is set with an alert before it is reached.
- ☐ Someone checks the cost against the value measure each month for the first three months.

8 · Operations

- ☐ Monitoring shows whether it is working: volume, errors, flags, cost.
- ☐ A runbook says how to restart it, roll it back, and switch it off.
- ☐ Changes go through a known process with the evaluation run first.
- ☐ Users were trained, and know who to call.

9 · The decision

Item	Answer
Exit criteria met?	
Open items, each with an owner and date	
Go / no go / go with conditions	
Decided by	
Date	
Review date (30 to 90 days after go-live)	