

What does it take to get an AI agent into production?

Clear limits on what the agent can do, tests on real examples, people reviewing the decisions that need judgment, monitoring after release, and a team that owns it.

By Allan Carroll, Managing Partner & CTO · Published September 23, 2026 · strataedge.ai

To get an AI agent into production, you decide exactly what it should produce, give it access to the right data and systems, test it on real examples, have people review the decisions that need judgment, monitor it after release, and put a team in charge of running it. A demonstration usually covers the first two or three steps. The rest takes longer. For Boston Consulting Group (BCG), StrataEdge and two engineers built a working AI agent in four weeks, and it was in production three months after we started.

What the agent can read and change

An agent reads information, produces something, such as a draft letter or a filled-in form, and sometimes takes an action in another system. Write each of those down: the records it can read, the systems it can change, and the actions that need a person's approval first. For example, an agent can be allowed to draft a customer letter while a person approves sending it.

Give the agent only the access it needs, approve that access through your normal process, and record every action it takes.

Where you can, have the agent fill in defined fields instead of writing free text. Ordinary code can then check the fields, and reviewers can compare them with the source. The agent we built for BCG uses structured outputs and has human review built in.

Testing before release

Collect examples with accepted answers, including incomplete and unusual ones. Before release, agree on the result you need, such as the share of results reviewers accept without changes. Run the same tests every time you change the model, the prompt, or the business rules. A model that passes in a demonstration can fail on inputs it has not seen, and a new model version can change results on cases that used to pass.

Which results people review

Decide which results go straight through and which a person reviews. Send a reviewer the results that fail validation, files with missing data, and high-stakes decisions, and show the evidence behind each result. Record what the reviewer decided and why. Those records show where the agent makes mistakes, and you can add them to your tests.

Plan the reviewers' time. If the agent sends people more work than they can handle, files wait in the review queue. Track that queue from the first week.

Monitoring and maintenance

Inputs change after release. New document layouts arrive, upstream systems rename fields, and people use the agent in ways nobody tested. Track how often reviewers correct it, the number of exceptions, processing time, and cost per item. Decide who fixes the agent when those numbers get worse.

Budget for maintenance from the start. Business rules change, source systems change, and model providers retire versions. Test each change against your accepted examples before it reaches production.

Who runs it

Someone in the business has to decide what the agent should do, and a team has to run it day to day. Train the people who review its work on the records and exceptions they will actually see. Give your engineers the tests, the monitoring, and the documentation they need to change it without us.

We build AI agents and automation, and we prepare client teams to use and maintain what we build. See our [typical engagements](#) or read the [BCG case](#).

[Talk with us about taking an agent to production](#)