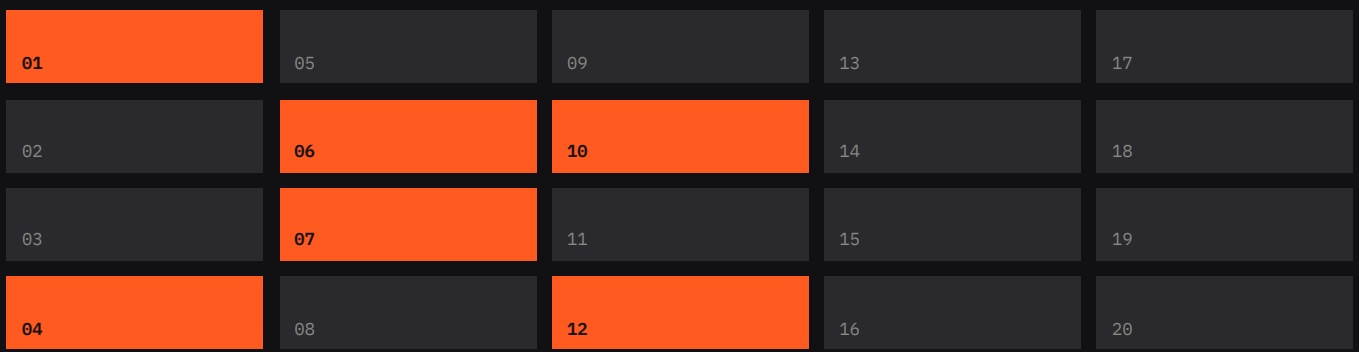
20 QUESTIONS · 5 BLOCKS · 6 BLOCKERS

Before you launch your AI agent.

The questions that separate a demo that impresses from an agent that can work in production.



Blocker: if the answer is "no," it doesn't launch

Check question

A demo works with the easy case. Production is everything else.

Building an agent that wows in a demo takes days. Getting it into real work is something else: the edge cases, the data that doesn't add up, someone trying to trick it, a cost that spirals, or a model change that alters its personality.

These are the questions I ask myself before I green-light an agent. I start with the challenge, not the tool. It isn't a standard or a certification: it's practical judgment, so you have the right conversation in time, not once the agent is already in production.

01

Answer it as a team

Business, technology, legal and whoever will run it, in the same room. Where you disagree, that's where the work is.

02

Be literal

Check what exists today, not what's planned. "We'll do it" counts as "no."

03

Score and decide

Yes scores 2, partly 1, no 0. Page 08 tells you what to do with the result.

■□□

Yes · Partly · No

One mark per question.

BLOCKER

Six knockout questions

A "no" on any of them stops the launch, whatever the rest adds up to.

▲

Red flag

The phrase that, if you hear it in the meeting, tells you the real answer is "no."

CHECKLIST MAP

01 **Use case & value**
01 02 03 04
PAGE 03

02 **Data & knowledge**
05 06 07 08
PAGE 04

03 **Agent design**
09 10 11 12
PAGE 05

04 **Risk, security & compliance**
13 14 15 16
PAGE 06

05 **Operations & measurement**
17 18 19 20
PAGE 07

01 Use case & value

Before talking about models: what problem it solves, for whom, and what it's really worth.

01

BLOCKER

Can you describe in one sentence what task the agent does, for whom, and when it considers the task done?

- [ ] Yes
- [ ] Partly
- [ ] No

WHY IT MATTERS

A scope that doesn't fit in one sentence can't be evaluated, bounded or defended.

▲ RED FLAG

"It'll help us with all things customer service."

02

Have you measured the baseline of the current process: time, cost and errors?

- [ ] Yes
- [ ] Partly
- [ ] No

WHY IT MATTERS

Without the "before," any improvement is an opinion and the project gets defended with anecdotes.

▲ RED FLAG

Improvement will be measured by the team's impressions.

03

Does the business case subtract human review, errors, handoffs and maintenance?

- [ ] Yes
- [ ] Partly
- [ ] No

WHY IT MATTERS

Gross savings aren't savings. What doesn't make it into the spreadsheet shows up in production.

▲ RED FLAG

The spreadsheet only has a savings column.

04

BLOCKER

Is there a business owner who decides, prioritizes and is accountable for the result?

- [ ] Yes
- [ ] Partly
- [ ] No

WHY IT MATTERS

Without a business owner, the agent becomes an eternal pilot or an IT toy.

▲ RED FLAG

Tech owns it "because it's an AI thing."

BLOCK SCORE ____ / 8

If there are noes here, you don't have an agent: you have a demo looking for an owner.

02 Data & knowledge

The agent answers with what it reads. If what it reads is wrong, it will act wrong, and with total confidence.

05 Does the agent draw on specific, up-to-date sources, with someone responsible for maintaining them?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

Its answers will never be better than its sources.

▲ RED FLAG

“Let’s give it access to the whole intranet and let it search.”

06 BLOCKER Do you know what personal or confidential data it will touch, and on what legal basis?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

Privacy sorted out late is the fastest way to stop an entire project.

▲ RED FLAG

Nobody has talked to the data protection officer yet.

07 BLOCKER Do you have a test set with real cases, including the hard and the rare ones?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

Without evaluation you don’t know whether one version is better than the last. You just have a hunch.

▲ RED FLAG

It gets tested by eye with five questions that always go well.

08 Have you defined what a correct answer is, and who judges it?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

If the criteria aren’t written down, every reviewer scores differently and there’s no way to improve.

▲ RED FLAG

Two people on the team can’t agree on the same answer.

BLOCK SCORE ____ / 8 If there are noes here, the agent will sound convincing and be wrong with total conviction.

03 Agent design

An assistant answers; an agent acts. The scope of what it can do is the size of its risk.

09 Are the actions it can take, and the ones it’s forbidden from taking, written down?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

Whatever isn’t bounded, sooner or later it will try.

▲ RED FLAG

“Let it do whatever it takes to solve it.”

10 BLOCKER Does it have only the minimum permissions its task needs, and not an administrator’s?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

Least privilege: if something fails or it gets manipulated, the damage stays contained.

▲ RED FLAG

It runs on the service account of someone with access to everything.

11 Do irreversible actions (payments, sends, deletions) require a person to confirm?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

What can’t be undone shouldn’t depend on a model alone.

▲ RED FLAG

It can email customers without anyone seeing the message first.

12 BLOCKER Can it say “I don’t know” and hand the case to a person, with full context?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

A good handoff beats a made-up answer. And the customer doesn’t have to repeat their story.

▲ RED FLAG

It always answers something, even when it has no information.

BLOCK SCORE ____ / 8

If there are noes here, the question isn’t whether it will get something wrong, but how much damage it can do when it does.

04 Risk, security & compliance

What it reads may try to manipulate it, and what it does has to be explainable.

13 Have you tested what it does when a document, an email or a web page gives it instructions?

WHY IT MATTERS

Prompt injection is a real risk for any agent that reads third-party content.

▲ RED FLAG

Nobody on the team has tried to trick it on purpose.

☐ Yes ☐ Partly ☐ No

14 Have you classified the risk of the use case, including what the EU AI Act requires?

WHY IT MATTERS

Bringing in legal at the end usually means rework. At the start, it just means designing well.

▲ RED FLAG

“Legal will review it before it goes live.”

☐ Yes ☐ Partly ☐ No

15 Do users know they’re talking to an AI?

WHY IT MATTERS

It’s a transparency obligation in many cases and, above all, a matter of trust.

▲ RED FLAG

The agent introduces itself with a human name and photo.

☐ Yes ☐ Partly ☐ No

16 Is there a record of what it did, with which data, and why?

WHY IT MATTERS

Without traceability there’s no audit, no debugging and no way to improve.

▲ RED FLAG

Only the final answer is stored, not the steps.

☐ Yes ☐ Partly ☐ No

BLOCK SCORE ____ / 8 If there are noes here, the risk isn’t technical: it’s reputational and legal.

05 Operations & measurement

Launch isn't the end of the project: it's the start of the real work.

17 Do you measure quality, cost per task and usage from day one, with alert thresholds?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

Quality shifts with real-world use. What isn't measured degrades silently.

▲ RED FLAG

The dashboard will get built "once there's traffic."

18 Do you have a plan B if the agent, or the model provider, fails?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

A business process can't depend on an external service always being up.

▲ RED FLAG

If the agent goes down, the process stops.

19 Do you know how you'll validate every change to the model, instructions or data before releasing it?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

An improvement in one place can break another. Your test set is the safety net.

▲ RED FLAG

Changes go straight to production.

20 Is there budget and a team to keep it alive after launch?

- ☐ Yes
- ☐ Partly
- ☐ No

WHY IT MATTERS

An agent in production needs ongoing care: data, tests, costs and users.

▲ RED FLAG

The project ends on launch day.

BLOCK SCORE ____ / 8 If there are noes here, it will work on demo day and degrade without anyone noticing.

Add up, draw and decide.

BLOCK	YES · ×2	PARTLY · ×1	NO · ×0	POINTS
01 Use case & value				/ 8
02 Data & knowledge				/ 8
03 Agent design				/ 8
04 Risk, security & compliance				/ 8
05 Operations & measurement				/ 8
Total				/ 40

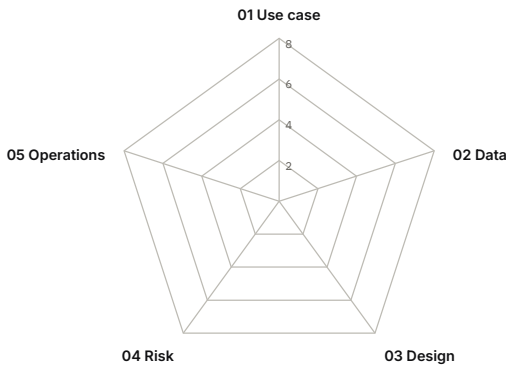
FIRST, THE BLOCKERS

Review questions 01, 04, 06, 07, 10 and 12. If any of them is a "no," the result is **don't launch yet**, whatever the total.

- ☐ 01
- ☐ 04
- ☐ 06
- ☐ 07
- ☐ 10
- ☐ 12

Check off each blocker you pass.

DRAW YOUR PROFILE



Mark each block's points (0–8) and connect them. The dips inward are your next conversation.

WHAT YOUR TOTAL MEANS

0 — 24 — 34 — 40

0–23 · Still a demo
Go back to the use case and the data. Building now means paying twice.

24–33 · Bounded pilot
Internal users, human review of everything and a review date.

34–40 · Controlled production
With metrics, thresholds and a plan B.

The thresholds are my practical judgment, not an industry standard. Use them to structure the decision, not to replace it.

If any question made you uncomfortable, it's doing its job.

TOOL · 3 MINUTES

Is your company ready for AI?

Twelve questions: your maturity level, your bottleneck and a 90-day plan.

rogergrossi.com/en/tools/ai-readiness-test

TOOL

Does your AI agent pay off?

The real net savings, after review, errors, handoffs and maintenance.

rogergrossi.com/en/tools/ai-agent-roi-calculator

CONVERSATION

Shall we talk?

If you're about to launch an agent and want a second opinion.

rogergrossi.com/en/contact

Roger Grossi · Business, data and AI strategist. I turn complex challenges into decisions that move the business.

© 2026 ROGER GROSSI · FREE TO SHARE WITH ATTRIBUTION.