

The eight gap scorecard

Score your own deployment before anyone sells you anything.

Eight patterns account for almost every agentic AI deployment that stalls. **Two of them are about the AI.** The other six are ordinary project work, and they are where most of these projects actually fail.

Score each one from 1 to 5, where **1 is "this is true of us right now"** and **5 is "we have this handled"**. Be honest; nobody sees this but you.

THE SIX ORDINARY ONES

1. Sponsorship UNIVERSAL	1 • No named executive owner. Decisions are slow. Junior teams without the full picture are driving the work, and trade offs get escalated but never settled.	5 • One named executive owns the outcome, stays current on progress, and manages risk before it becomes a delay.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
2. Requirements UNIVERSAL	1 • Nothing is locked. No signed off process documentation, no real sample data, and the business and the subject matter experts describe the process differently. Rules and measures arrive late, so scope drifts.	5 • Requirements and sample data are locked before the build starts. Edge cases are mapped up front, and the undocumented knowledge in people's heads has been written down.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
3. Infrastructure UNIVERSAL	1 • Credentials, environments and test setups take weeks, not days. The project starts before the environment is ready and waits on access it should already have had.	5 • Infrastructure is specified once, completely, and stood up before any code. Cloud, on premise or hybrid is decided, and the test environment exists before the build.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
4. Focus UNIVERSAL	1 • Chasing every case instead of the vital few. Time goes into rare exceptions before the common path is live, and the team is tired of pursuing perfection.	5 • The high volume, high impact cases go first. The rest are scheduled deliberately, after something is running.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
5. Integration RPA LEGACY	1 • It can read your systems but it cannot reliably write into them. Every system of record has to be read and written accurately and in real time, and most stacks never cross that line.	5 • Validated write paths into the systems you already run, including the legacy portals, with no rip and replace.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
6. Operational RPA LEGACY	1 • It works on clean test data. It has never been through a messy month: poor scans, partial documents, supplier variations, real volume.	5 • Accuracy holds at production volume and messiness, with deterministic checks where the answer must never vary, and a real exception path.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

THE TWO THAT ARE NEW WITH AGENTS

7. Trust NEW WITH AGENTS	1 • You cannot say in one sentence where your data goes. Public models cannot meet your accuracy bar, and compliance, legal or security are blocking production.	5 • Governance from day one. Data security, privacy and residency are settled, guardrails and policy controls are in place, and you can trace any decision.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
8. Accountability NEW WITH AGENTS	1 • Nobody owns drift. The vendor blames the integrator, the integrator blames the data, performance degrades quietly after go live, and running cost is opaque.	5 • One owner after go live, monitoring drift over time, accountable for the outcome and for what it costs to run.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Reading your score

- **Any 1 or 2 on items 1 to 6:** your problem is not the model. Fix these before you evaluate another vendor, because no vendor can fix them for you.
- **Any 1 or 2 on items 7 or 8:** these are the genuinely new ones. They are also the ones a serious partner should be able to answer immediately.
- **Three or more at 1 or 2 in total:** the deployment will stall. That is not a prediction, it is what these eight patterns describe.

Before you let anything write into a production system

Three questions. If you can answer each in one sentence, you are ready. **If any of the three needs a paragraph, keep that surface read only for now.**

1. **What may it touch?**
2. **Who approves?**
3. **How do you reverse a bad run?**

Where to start, when you are ready

Score each candidate workflow on three things. The first deployment usually becomes obvious.

- **Volume.** Hundreds of decisions a day, not dozens a quarter.
- **Pain.** Expensive to be wrong: margin, customers or compliance.
- **Data.** Already digital, in one place, with a clean system of record.

One question for any vendor, including us

Accuracy is quoted per field, but a document posts as a whole. If twelve fields must all be correct and each is right 98% of the time, about 78% of documents come out fully correct. At 99% a field, about 88%.

So the number to ask for is not accuracy. It is:

What share of documents completed last month with nobody touching them?



hachiai.com/allin2026

The scorecard, the playbook, the paper Lisa Hyde co-wrote, and the recording of the session once it is available. Nothing to fill in to read the card.