

The AI Readiness Playbook

Find where AI pays off in your business, before you spend a dollar building it.

Density Labs · densitylabs.io · 2026 edition

95% of AI pilots never reach production (MIT NANDA, 2025). Not because the technology fails, but because there was no plan that survived contact with the company's real workflows, data, and team. This playbook is the plan. It is the same three step loop we run in paid engagements, Map, Prove, Deploy, written so you can run the first pass yourself in one working week.

Step 1: Map (days 1 to 3)

AI creates value in workflows, not in mission statements. Your first job is to find the workflow where it pays.

Make the candidate list

List every repeatable workflow that eats paid hours: quoting, scheduling, document handling, support replies, reporting, data entry, follow ups. Ask each team lead one question: "which task would you never miss if it disappeared?"

Watch the work

For your top three candidates, sit with the person doing the work for thirty minutes. Not a meeting about the work. The work. Write down every step, every exception, every system touched. The exceptions are the real spec.

Score every candidate

The scoring card. Rate each workflow 1 to 5 on:

- **Hours:** how many paid hours per week does it consume?
- **Frequency:** does the decision repeat often enough for automation to learn and pay off?
- **Data:** do examples of the work exist digitally, and are they findable?
- **Error tolerance:** if the AI is wrong 1 time in 10, does a human catch it cheaply?
- **Ownership:** is there one person who owns the output and can say what "good" looks like?

Multiply Hours by Frequency, then subtract points where Data, Error tolerance, or Ownership scored below 3. The highest total is your first project. Anything scoring under 3 on Data is not ready, fix the data first.

Estimate ROI and TCO honestly

ROI: hours saved per week, times loaded hourly cost, times 50 weeks. TCO: build cost plus model usage plus the human review time that stays, plus maintenance (assume 20% of build cost per year). If projected ROI is not at least 3x TCO, pick the next candidate.

Step 2: Prove (days 4 to 5, then ongoing)

Never trust a demo. Demos are the reason for the 95% statistic. Trust a measured pass rate.

Build a golden set

Collect 20 to 50 real, past examples of the workflow with known correct outputs: real tickets with the reply that was actually sent, real documents with the data that was actually extracted. Label them yourself or with the workflow owner. This is the single highest leverage artifact in the whole project.

Measure before you automate

Run any candidate tool or prototype against the golden set. Count exact passes, near misses, and failures. Write the number down. If a vendor cannot run against your golden set, that is your answer about the vendor.

The threshold rule. Below 70% pass: not viable yet, improve the data or narrow the scope. 70 to 90%: viable with a human approving every output. Above 90% on a mature golden set: candidates for limited autonomy on the lowest risk slice.

Step 3: Deploy (weeks 2 to 12)

Autonomy is earned in stages, never granted on day one.

1. **Sandbox.** The system runs on real inputs, humans see its output, nothing leaves the building. Keep measuring against the golden set.
2. **Assist.** The system drafts, a human approves every send. First response time drops while accuracy stays visible.
3. **Limited autonomy.** Only the categories with proven pass rates run unattended, with sampling and an escalation path. Everything else stays assisted.

Log every input, output, and override. The overrides are next month's training data. And decide before launch who owns the system when it misbehaves, "nobody owned it" is the most common cause of quiet death after a promising pilot.

The four red flags when buying help

- Nobody asks to watch the actual work; they scope from a questionnaire.
- The deliverable is a slide deck, not a plan your team can execute alone.
- No golden set, no measured accuracy, only demos.
- The recommendation is always the tool they resell. An honest review sometimes says: build nothing.

Where to start, by role

You are	Start with	Your trap to avoid
A technical leader (CTO, VP Eng)	The golden set. Your team can build one this week, and it converts every AI debate into a measurement.	Building the flashy feature first. Make the first use case boring on purpose.
A non technical founder or owner	The scoring card. You know where the hours go better than anyone.	Buying a tool before mapping the workflow. Tools are step three, not step one.
An individual contributor	Map your own workflow and bring the scoring card to your manager with numbers attached.	Automating in the shadows. Unowned AI gets banned the first time it embarrasses someone.

When to stop self serving

This playbook takes you through the first honest map and the first measured prototype. Call in help when: the top workflow crosses systems nobody fully understands, the golden set says 70 to 90% and closing the gap needs real engineering, or the plan needs numbers finance will sign. That is the line where our paid engagement begins: the **AI Readiness Assessment**, \$2,500, two weeks, a senior engineer runs this entire loop on your operation and hands you the written plan. densitylabs.io/ai-readiness-assessment

Density Labs · The AI Engineering Partner for the US mid market · densitylabs.io · hello@densitylabs.io
96% client retention, Clutch verified. You may share this document freely with your team.