

✦ TEMPLATE · TEMPLATES

# AI MVP Planning

Plan an AI minimum viable product that tests the riskiest assumption first: users, the core job, quality bar, data, scope cuts, build plan and the decision you will make at the end.

| FORMAT | LEVEL | READ TIME | UPDATED |
| --- | --- | --- | --- |
| Template | Beginner | Fill-in | Oct 2026 |

OVERVIEW

# What this template covers

AI MVPs fail differently from ordinary ones. The interface is rarely the risk; the risk is whether the model can do the job well enough, on your data, at a cost that works. This template forces those questions into the plan, so the MVP answers them instead of producing a polished demo that proves nothing.

## KEY TAKEAWAYS

- An MVP exists to answer a question. Write the question down first.
- Test model quality on real data in week one, before building the product around it.
- Define “good enough” as a measurable quality bar agreed with users.
- Cut scope by user journey, not by feature: one journey done well beats five half-built.
- Decide in advance what result means continue, pivot or stop.

## WHO IT’S FOR

- Founders and product managers
- Innovation and digital teams
- AI engineering leads
- Sponsors funding a first build

## HOW TO USE IT

Fill sections 01–03 before any build starts, ideally with two or three target users in the room; sections 04–05 become the build plan and section 06 records the decision. Every box is shown filled in with a worked example — a referral intake assistant for a fictional chain of eight therapy centers — so you can see the level of detail expected. Replace the examples with your own.

## CONTENTS

01 Users, job and riskiest assumption
02 Quality bar and success criteria
03 Data and feasibility check
04 Scope: in, out and later
05 Build plan and budget
06 Pilot results and decision

01 / HYPOTHESIS

# Users, job and riskiest assumption

Describe the person and the job before the product.

## Target user

01

A specific person: role, context, and what they use today. e.g. “Front-desk staff at a therapy centre, booking on phone and WhatsApp.”

WORKED EXAMPLE

Front-desk coordinators at a fictional chain of eight child-therapy centers. Each handles 25–40 new referrals a week arriving as WhatsApp photos, emailed PDFs and walk-in paper forms, and re-types them into the clinic’s booking system.

## The job to be done

02

“When [situation], I want to [motivation], so I can [outcome].”

WORKED EXAMPLE

When a new referral arrives, I want the child’s details, the referring doctor and the requested therapies captured correctly in our system, so I can call the family and book the first assessment the same day.

## Why AI, and why not a simpler tool

03

What part genuinely needs language understanding, generation or judgment?

WORKED EXAMPLE

Referrals come from 60+ doctors on different letterheads, often handwritten and in English or Tamil. A fixed form or template OCR breaks on that variety; extracting meaning from messy documents is the part that genuinely needs a language model.

## Riskiest assumption

04

The belief that, if wrong, kills the idea. e.g. “The model can extract fields from handwritten referral forms accurately enough.”

WORKED EXAMPLE

The model can extract the eight booking fields (child’s name, date of birth, guardian phone, referring doctor, diagnosis, therapies requested, urgency, notes) from photographed handwritten referrals accurately enough that checking is faster than typing.

### Rank your assumptions

| ASSUMPTION TYPE | QUESTION IT ANSWERS | HOW TO TEST CHEAPLY |
|---|---|---|
| **Desirability** | Do users want this enough to change habits? | Interviews, concierge test, clickable prototype |
| **Feasibility** | Can the model do the task well on our data? | Notebook test on a sample of real inputs |
| **Viability** | Does cost per task work for the business? | Price the test runs and project volume |
| **Trust** | Will users and reviewers accept the outputs? | Side-by-side review with users |

✦ 02 / QUALITY

# Quality bar and success criteria

Agree with users what “good enough to use” means — it is usually lower than perfect and higher than the demo.

### Quality bar

01

e.g. “Reviewers accept the draft with minor edits in most cases; no factual errors on dosage, pricing or dates.”

WORKED EXAMPLE

Coordinators accept the extracted record with at most two field edits in 8 of 10 referrals, and checking a record takes under 90 seconds, against about 6 minutes to type it today.

### Unacceptable failures

02

Errors that would end trust immediately, regardless of averages.

WORKED EXAMPLE

- A wrong guardian phone number saved without being flagged
- Two children’s details mixed in one record
- A field filled in that is not present on the referral

### MVP success criteria

03

Usage, task time, acceptance rate, willingness to pay — with the threshold for each.

WORKED EXAMPLE

- Used for at least 70% of new referrals at the two pilot centers
- Median intake time under 2 minutes per referral
- Acceptance rate (≤ 2 edits) of 80% or more
- Center managers say they would fund roll-out to all eight centers

✦ 03 / DATA

# Data and feasibility check

Run a quick feasibility test on real inputs before committing to the build.

### Data needed and where it comes from

01

Sources, formats, volumes, owners and consent or legal basis for personal data.

WORKED EXAMPLE

300 past referrals (photos and PDFs) from two centers, with the records coordinators actually typed as the expected output. Families consented to referral processing at registration; the sample is de-identified for the feasibility test and stays on India-region storage.

### Feasibility test plan

02

Sample size, who labels the expected outputs, which models or approaches you will compare.

WORKED EXAMPLE

- 100 referrals: 60 handwritten, 30 typed, 10 in Tamil; a senior coordinator checks the expected fields
- Compare two vision-capable models and one OCR + text-model pipeline
- Score field-level accuracy and time to review

### Feasibility test result

03

Completed after the test: what worked, what failed, and whether the quality bar looks reachable.

WORKED EXAMPLE

- Typed referrals 96% of fields correct; handwritten 84%, with phone numbers weakest (78%)
- Fix: show phone and date of birth beside a crop of the source image
- Verdict: quality bar looks reachable; proceed to the thin slice

✦ 04 / SCOPE

# Scope: in, out and later

Pick one end-to-end journey. Everything else goes to “later”.

| TYPICAL MVP OVER-REACH | A FOCUSED AI MVP |
|---|---|
| Several user types at once | One user type, one journey |
| Every channel from day one | One channel they already use |
| Custom model training | Existing models + good retrieval and prompts |
| Full admin and analytics suite | Logs and a simple review screen |

### The one journey in scope

01

Start to finish, in steps the user takes.

WORKED EXAMPLE

- Coordinator forwards a WhatsApp photo or PDF to the intake inbox
- Assistant extracts the eight fields and flags low-confidence ones
- Coordinator checks each field beside the image and saves it to the booking system, which creates the usual “call family” task

### Explicitly out of scope

02

WORKED EXAMPLE

Walk-in paper forms, automatic messages to families or doctors, and booking the assessment slot.

### Later (if the MVP succeeds)

03

WORKED EXAMPLE

- Roll-out to all eight centers, with walk-in scanning
- Suggest a therapist from the requested therapies

✦ 05 / PLAN

# Build plan and budget

Short, fixed time boxes keep an MVP honest.

01 Feasibility test → 02 Thin end-to-end slice → 03 Evaluation set + review → 04 Pilot with real users → 05 Decision

## Milestones

| MILESTONE | EXIT CRITERION | OWNER | TARGET DATE |
|---|---|---|---|
| **Feasibility test** | Quality bar reachable on sample | Product engineer | 23 Oct 2026 |
| **Thin slice** | One journey works end to end | Product engineer | 6 Nov 2026 |
| **Evaluation** | Golden set scored, failures understood | Senior coordinator | 13 Nov 2026 |
| **Pilot** | Users complete real tasks | Operations Director | 11 Dec 2026 |
| **Decision** | Continue / pivot / stop recorded | Operations Director | 18 Dec 2026 |

### Team and budget

01

People and time committed; build budget; model and hosting cost ceiling for the pilot.

WORKED EXAMPLE

One product engineer full time and one designer at 50% for six weeks; a senior coordinator at 20%. Build budget ₹9 lakh; model and hosting ceiling ₹15,000 for the pilot month.

✦ 06 / DECISION

# Pilot results and decision

Fill this in at the end of the pilot, against the criteria in section 02.

## Before deciding

- [ ] Results compared with success criteria set before the build
- [ ] Real users, not only the team, completed real tasks
- [ ] Cost per task measured at pilot volume
- [ ] Unacceptable failures counted and understood

### What we learned

01

About users, model quality, data and cost.

WORKED EXAMPLE

- Coordinators trusted the tool once the source crop sat beside each field
- Field accuracy 93% typed, 87% handwritten; median intake time fell from about 6 minutes to 1 min 40 s
- Cost per referral about ₹4 at pilot volume

### Decision

02

Continue to production / pivot (how) / stop — and why.

WORKED EXAMPLE

**Continue to production** for all eight centers, with walk-in scanning added first. Phone-number confirmation stays mandatory until handwritten accuracy is above 95%.

### Decided by — names and date

03

WORKED EXAMPLE

K. Raman, Operations Director · N. Bose, Clinical Lead · 18 Dec 2026

✦ GET IN TOUCH

# Want us to run this with your team?

We run short, fixed-scope AI MVPs that test feasibility on your real data first, then put a working slice in front of real users.

- ✦ Feasibility sprints
- ✦ AI MVP builds
- ✦ Pilot evaluation and go / no-go

BOOK AN AI STRATEGY SESSION →

EMAIL
hello@splendenslabs.com

WEB
splendenslabs.com

OFFICE
Sona Incubation Foundation, Salem