

# The AI Product Spec

A Hands-On Tutorial for Defining and Scoping Recommendation Systems

**Rohan Mitra**

Product Manager @ PhonePe (a Walmart co.)



**Madhu Guru** ✓  
@realmadhuguru

At @Google, we are moving from a writing-first culture to a building-first one.

Writing was a proxy for clear thinking, optimized for scarce eng resources and long dev cycles - you had to get it right before you built.

Now, when time to vibe-code prototype  $\approx$  time to write PRD, PMs can SHOW not tell. Role profiles are blurring, creativity and building are happening in parallel.

**"So, is the PRD dead?"**

# The "Vibes-Based" Spec: A Masterclass in Saying Nothing

## Example 1: Smart Returns Vague Goals & Wishes

---

- **Problem:** "Returns are a big problem."
- **Goal:** "Improve customer experience."

*Result: Ambiguity*

[PRD Example 1 - Before - Smart Returns](#)

## Example 2: Smart Replies Generic Specs

---

- **Goal:** "Improve engagement."
- **AI Behavior:** "Generate helpful replies."
- **Rollout:** "Start small, then ramp."

*Result: Wasted Sprints*

[PRD Example 2 - Before - Smart Replies](#)

# The 5-Pillar Framework



## 1. Opportunity Framing

Hypothesis & Strategy



## 2. Behavior Contract

Examples & Personality



## 3. Evaluation Plan

Golden Sets & Metrics



## 4. Boundaries & Scope

Non-Goals & Tradeoffs



## 5. Risks & Recovery

Kill Switches & Fallbacks

# Deep Dive 1: The Behavior Contract

THE TRADITIONAL SPEC

## Logic-Based

---

"System should generate helpful replies."

→**Result:** Engineer guesses what "helpful" means.

THE AI SPEC

## Example-Based

---

**GOOD:** (Actionable)

**BAD:** (Vague/Robotic)

**REJECT:** (Safety Violation)

# Deep Dive 2: The Evaluation Plan

## Step 1: Offline

Test: 500 labeled "Golden Set"  
examples

**F1 Score  $\geq 0.85$**

## Step 2: Human Review

Test: Manual review of 200  
responses

**Rating  $\geq 4/5$**

## Step 3: Online Guardrails

Test: Live traffic A/B test

**Blocked  $\leq +0.3\%$**



# Live Demo: Fixing "Smart Replies"

**Goal:** Turn "Generate helpful replies" into a decision-based Behavior Contract.

# The V1 Spec (Our Result)





## THE VIBE (BEFORE)

"Generate helpful replies."

---

No clear pass/fail criteria.

## THE V1 SPEC (AFTER)

-  **Proxy:** Adoption rate  $\geq 25\%$
-  **Guardrail:** Blocked rate  $\leq +0.3\text{pp}$
-  **Tradeoff:** +300ms latency allowed
-  **Behavior:** GOOD="Sending by 4pm"

# Your Framework for AI Product Leadership



## 1. Navigate Politics

Turn "First Principles rhetoric" into real, explicit tradeoffs.



## 2. Clarify Chaos

Turn "Vibe Coding" into a Behavior Contract.



## 3. Build Skills

Practice Scrappiness, Low Ego, & Clear Communication.

# Thank You

**Rohan Mitra**

[linkedin.com/in/rohanmitra](https://www.linkedin.com/in/rohanmitra) | [@irohanmitra](https://twitter.com/irohanmitra)