

Direction, Discovery, and Real Evals Define the PM Edge

PM Daily Digest

2026-06-12

Direction, Discovery, and Real Evals Define the PM Edge

By PM Daily Digest • June 12, 2026

This brief covers a sharper mental model for AI-era product roles, a practical discovery and evals playbook, Meesho's customer-led pivots, and fresh PM hiring and job-search signals.

Big Ideas

- **AI expands roles before it erases them.** Sachin Rekhi argues product, design, and engineering are not collapsing into one “AI builder” blob; each circle is expanding. If AI doubles engineering throughput, the leverage shifts toward clearer direction on what to build and why. Role blending removes coordination bottlenecks, but differentiated strategy, design, and frontier engineering still need specialists [1, 2]. **Why it matters:** PM value rises with framing, prioritization, and decision quality. **Apply it:** let PMs prototype and designers ship polish, but keep explicit ownership for strategy choices and quality bars.
- **Good strategy is built before the strategy deck.** Scott Belsky's point: teams often solidify strategy only when they need to present it. Exploring “the edges that may someday become the center” and running experiments early makes bolder decisions easier later [3]. **Why it matters:** faster execution exposes weak assumptions faster. **Apply it:** keep a small queue of edge bets and socialize what you learn before quarterly or annual planning.

Tactical Playbook

- **Use reverse demos for discovery.** Musubi starts onboarding by having customers walk through their current moderation system and show what is

failing - often false positives or systems that cannot adapt to new attacks [4]. From there, the team proposes a fit-for-problem mix from a reusable toolkit rather than defaulting to bespoke work [4]. **Why it matters:** you get grounded in real failure modes, not abstract requirements. **Apply it:** (1) ask customers to show the live workflow, (2) capture where it breaks, (3) map those failures to capabilities, (4) generalize only after repeated demand across customers [4].

- **Treat evals as operating work, not vocabulary.** In OpenAI PM hiring conversations, candidates stood out by running real evals, writing rubrics for failures, and measuring improvement on actual builds - not by talking about evals abstractly [5]. Musubi pushes the same discipline into the product with customer-managed golden sets, automated error analysis, and human review to avoid overfitting [4]. A solo builder of PasteFlow made the same point from another angle: prompting was maybe 10% of the work; the rest was PRDs, edge cases, scope control, and defect triage [6]. **Apply it:** start with one failing workflow, define a golden set or rubric, review false positives and negatives, and keep a human decision-maker in the loop.

Case Studies & Lessons

- **Meesho found fit by observing real behavior, then changing segments hard.** The team first listened only to sellers and missed the consumer side; when they pushed the app to buyers, consumers called it the “worst of both worlds” versus malls or e-commerce [7]. Sitting in shops revealed the real behavior: many merchants were already “online” through WhatsApp groups, which functioned as the storefront [7]. Meesho then focused on online-native sellers, launched Meesho Supply, and saw organic usage double month over month with high retention [7]. Later, even with a business serving 10 million sellers, the company committed to a direct consumer app after fresh field research showed many assumptions about app-download friction no longer held [7].

“Be very rigid with your problem and be very flexible with your solution.” [7]

Takeaway: stay close enough to customers to see behaviors competitors miss, then be willing to re-segment or abandon a successful channel when the long-term user reality changes [7].

Career Corner

- **The hiring signal is PM depth plus proof of work.** One cited benchmark put OpenAI’s median PM compensation at \$860K [5]. More useful than the number: four OpenAI PM conversations emphasized deep PM fundamentals, shipping something real, and being able to explain evals

from firsthand experience [5]. Common misses were shallow AI familiarity, eval jargon without actual evals, and repos nobody uses [5]. **Apply it:** build one small product, let people use it, track where it breaks, and document how you measured improvements [5].

- **Practitioner job-search advice is getting more tactical.** In community discussion, PMs recommended optimizing resumes for ATS keywords, applying on company career pages instead of Easy Apply, and posting AI projects on LinkedIn to attract recruiter outreach [8]. Another shared constraint: regulated industries may screen hard for direct domain experience, so adjacent sectors can be a more realistic bridge [9].

Tools & Resources

- **Lightweight proof-of-work stack:** Claude Code plus Lovable or Replit were recommended as fast ways to build public AI projects that demonstrate PM judgment, not just prompt fluency [8, 5].
- **Job-search helpers:** BuiltIn, Motion Recruitment, Oliver James, and the HideJobs plugin were specifically recommended for filtering opportunities and reducing LinkedIn noise [9].

Sources

1. X post by @sherifmansour
2. X post by @sachinrekhi
3. X post by @scottbelsky
4. Beyond Black Box Scores: How Musubi Trains Custom AI for Trust and Safety Teams
5. substack
6. r/ProductManagement post by u/devbyroman
7. How Meesho Became India's Biggest Shopping App
8. r/ProductManagement comment by u/productman26
9. r/ProductManagement comment by u/ant3k