

Handoffs Collapse, Evals Rise, and PM Work Gets More Technical

PM Daily Digest

2026-05-29

Handoffs Collapse, Evals Rise, and PM Work Gets More Technical

By PM Daily Digest • May 29, 2026

This brief covers the collapse of traditional PM-design-engineering handoffs, practical patterns for building eval-driven AI products, and the career shift toward more technical, founder-like product work.

Big Ideas

- **AI is collapsing the PM-design-engineering relay race.** At OpenAI, PMs write markdown PRDs that Codex turns into shipped PRs; designers ship full UIs with instrumented noop backends; engineers focus on constraints, review agents, and build-failing tests in a shared repo [1]. Gokul Rajaram’s “pod-of-one builder” describes the same direction: AI compresses execution, so the scarce skill becomes judgment—choosing the right problem and spotting mediocre output [2]. **Why it matters:** coordination work is shrinking; product leverage is moving toward taste, validation design, and system constraints.
- **The interface shift is from task-based work to agent orchestration.** Emma from Resonant says PMs are becoming “agent orchestrators” who train agents over time so autonomous workflows do not turn into “AI slop” [3]. Scott Belsky makes the product-level version of the same point: replace task-based workflows with ask-based workflows at the OS level, while keeping underlying models swappable beneath the UI [4]. **Apply it:** encode principles, examples, and guardrails in reusable artifacts agents can act on.

Tactical Playbook

- Use a painted door before you build the backend.

1. Ship a real UI with a noop backend [1].
2. Instrument clicks and flows to see where demand is real [1].
3. Build APIs only where behavior justifies the investment.

This gives product and design real evidence instead of speculative prioritization.

- **Start AI products with evals, not giant prompt boxes.** Lorikeet is moving from “give us your SOP” toward first defining what good and bad outcomes look like, then having a Coach agent generate SOPs, guardrails, and test cases [5]. When the model is close but missing one fact, they use “resolution in the loop”: the AI pauses for targeted human input instead of escalating the whole ticket [5]. **Apply it:** define failure cases, escalation rules, and knowledge gaps before you tune prompts.

Case Studies & Lessons

- **Lorikeet followed customer pull, not founder intuition.** The team spent months on reflection tools and information dashboards before a healthcare startup made the real job explicit: help clear the support inbox [5, 6]. Their first prototype was a command-line workflow using real customer CSVs, which let them iterate quickly on real data [5]. Today the product runs a ticket-handling Concierge agent plus a Coach agent for configuration and improvement [5, 6]. At scaled customers, human average handle time went **up** because people were now spending more time on the hardest tickets [5]. **Takeaway:** good AI automation can increase the value of human work rather than simply reduce it.
- **AI can compress enterprise discovery into traceable artifacts.** In the ACNA example, a messy **\$2.8B** settlement with complex eligibility rules became a “First Pass” pre-validation layer that checks completeness before claims reach reviewers [7]. The workflow then turned those rules into decomposed user stories, test cases, and a living playbook [7]. **Takeaway:** in dense legal or operational domains, AI is most useful when it makes ambiguity auditable.

Career Corner

- **Founder skills are becoming PM skills.** Emma argues PMs are increasingly founder-like and well positioned to start companies [3]. The gaps she highlights are practical: getting from prototype to credible demo, understanding fundraising dynamics, and choosing between sales-led and product-led growth [3]. Her view: mildly technical PMs can now get to a real customer demo with tools like Lovable, and sales-led motions may be an easier way to get early traction than pure growth hacking [3]. **Apply it:** build one demo yourself and be ready to talk about distribution, not just features.
- **AI PM interviews are getting more technical.** Exponent reports that Meta’s AI product sense round combines a traditional 30-minute

product case with 30 minutes of live prototyping in Llama, followed by questions on token efficiency, latency, retrieval, and compute trade-offs [8].

“The one thing I told them to remember above all in the interview was: tell them what you’ve learned about users.” [9]

Apply it: prepare both a product narrative and a technical one: what you would build, what data it needs, where performance trade-offs show up, and use real data sources when possible [8].

Tools & Resources

- **Product Playbook pattern for AI discovery.** The workflow moves from freeform idea exploration to vision, personas, JTBD, event flows, user stories, test cases, and lean canvas outputs, then stores them in a living “Product Playbook” that teams can revisit as markets change [7]. Worth exploring if your team keeps losing context across meetings and stale docs.
- **Lovable for fast customer demos.** Emma cites it as a practical bridge from rough prototype toward something a mildly technical PM can show customers for feedback, even if it is not yet production-grade [3].

Sources

1. substack
2. X post by @sachinrekhi
3. Lessons from Shifting PM to Founder | ProductTank London
4. X post by @scottbelsky
5. Building Lorikeet: How AI Humility and a Dual-Agent Architecture Are Redefining Customer Support
6. X post by @ttorres
7. Building Products That Achieve Org Outcomes with AI & Human Experts | Rezoomex
8. What REALLY Happens in Meta’s AI Product Sense Round
9. X post by @paulg