

Live-Code Decisions, Experimentation Culture, and the New AI PM Playbook

PM Daily Digest

2026-06-04

Live-Code Decisions, Experimentation Culture, and the New AI PM Playbook

By PM Daily Digest • June 4, 2026

Teams are moving more product decisions into working code, leaders are re-learning how to make experimentation stick, and PMs have new guidance on AI evaluation, stakeholder influence, and frontier-lab career prep.

Big Ideas

- **Product quality decisions are moving from mocks to live code.** Anthropic's Head of Design said quality gates have shifted from PRDs, mocks, and Figma into working code, with small **3-5 person pods** making decisions and releasing internally before expanding externally based on real adoption. **Why it matters:** PMs can evaluate actual behavior earlier, not just intent. **Apply it:** replace some review cycles with working prototypes and internal dogfooding, then judge success on adoption, retention, and revenue rather than token counts alone. [1]
- **Experimentation only sticks when leadership turns it into culture.** David Bland warns that experimentation becomes theater when teams run tests only to justify a launch they already want. Monica Lewis adds that leaders need to normalize mistakes, share early thinking, and create discovery time, or teams revert to old habits. **Why it matters:** process without leadership behavior rarely lasts. **Apply it:** point experiments at real high-uncertainty opportunities, review what was learned, and have leaders keep modeling the behavior publicly. [2]

It was in our bloodstream, but it wasn't in our DNA [2]

Tactical Playbook

1. Use **signal prep before high-stakes meetings**.
 - Answer three questions: **What do I need from this room? What is my one-line recommendation? What will people repeat without me?** [3]
 - Lead with the destination, not a long backstory. That is especially useful for PMs who default to detail to prove credibility and then get labeled non-strategic. [3]
 - **Why it matters:** it shifts you from giving updates to leading a decision. **Apply it:** do a short prep pass before roadmap reviews, exec syncs, and stakeholder negotiations. In one coaching case, this shift changed how a Head of Product was perceived within 2-3 months. [3]
2. For **AI products, choose metrics by task, not convenience**.
 - Define the task precisely first. Accuracy can hide failure in imbalanced problems; **F1** is more useful for fraud, credit risk, and document classification. [4]
 - Use **BLEU** when the main risk is saying the wrong thing, **ROUGE** when the main risk is leaving out the right thing, **Exact Match + token F1** for extractive QA, and **perplexity** for model selection rather than production health. [4]
 - **Why it matters:** a single metric is easy to game. **Apply it:** track at least two complementary metrics and pair them with human evaluation before shipping. [4]

Case Studies & Lessons

- **Claude Code's operating model:** Anthropic said the product made **\$2.5B in its first year** and reached about **51% of the coding market**. The team ships through small pods, supports broad shipping authority with code review/CI/testing, and expands from internal use to external rollout after seeing real adoption. Enterprise growth has also been bottom-up, with developers becoming internal advocates and teams building connectors and tooling around the product. **Lesson:** speed scales when governance and shared infrastructure scale with it. **Apply it:** ship smaller internal-first releases and invest early in the tooling that makes adoption easier across a team. [1]
- **OpenAI PM leverage through synthesis:** Abhi Muchhal's setup includes a daily Slack triage for blockers and deadlines, a self-updating market dashboard pulling from **7-8 sources**, and a weekly stakeholder update drafted from Slack, Drive, Notion, and dashboards. **Lesson:** the highest-value PM automations are often synthesis workflows, not generic note-taking. **Apply it:** start with one recurring digest or dashboard that pulls from multiple systems but still keeps a human review step before anything goes out. [5]

- **Copilot's early signal:** Mario Rodriguez said initial acceptance rates were only **20-30%**, yet the product still created major value when suggestions were useful. **Lesson:** a weak surface metric can still mask strong product value. **Apply it:** pair AI interaction metrics with downstream outcome metrics and keep the learning loop fast. [2]

Career Corner

- **Frontier-lab PM hiring still starts with PM fundamentals.** Aakash Gupta's reporting says strong candidates show structured thinking, analytical decision-making, and communication under ambiguity, then prove AI fluency by building a real API-based project and speaking the language of evals: capability, baseline, and improvement criteria. **Why it matters:** tool familiarity alone is not the bar. **Apply it:** bring one real project you built and be ready to explain how you measured whether it improved. [5]
- **For strategy and design interviews, rehearse a default structure.** One practical format is **context, goal, user, constraints, options, tradeoffs, decision.** Candidates also recommended practicing on a company's top products and starting with clarifying why-questions. **Why it matters:** these interviews reward structured thinking under pressure. **Apply it:** practice aloud with a timer until the framework feels automatic. [6, 7, 8]

Tools & Resources

- AI Product Evaluation Framework, Simply Explained — a useful reference for matching NLP tasks to the right metric mix before shipping. [4]
- How to lead when you don't fit in — worth bookmarking for the CALM leadership model and the signal prep exercise. [3]

Sources

1. Anthropic Head of Design on How Claude Code Hit \$2.5B in Year One | Meaghan Choi | E298
2. Episode 270: How Experimentation Becomes Culture
3. How to lead when you don't fit in - Dave Martin (CPO, Fractional)
4. AI Product Evaluation Framework, Simply Explained
5. How to Use Codex Like an OpenAI PM | Abhi Muchhal, PM OpenAI (ex-Meta and Nubank)
6. r/ProductManagement comment by u/Born_Read121
7. r/ProductManagement comment by u/satishmummareddy
8. r/ProductManagement comment by u/titdaer