AI is changing how PMs are hired, and what counts as evidence

PM Daily Digest

2026-10-30

AI is changing how PMs are hired, and what counts as evidence

By PM Daily Digest • October 30, 2026

Take-home assignments and audit interviews test the judgment AI can't supply. Ravi Mehta and Hiten Shah separate observed, believed and tested evidence. A redesign backlash thread shows how to keep a rollback debate grounded in evidence.

Hiring now tests the judgment AI can't supply

Aakash Gupta estimates that 40–45% of PM job seekers face at least one take-home or past-project presentation, and that 16–31% of PM hiring processes ask for one. His sources are his own interviews and cohorts of 160 job seekers [1]. Anthropic's candidate rules say not to use Claude unless the company allows it [1]. According to Gupta, graders ask three things: does the work beat the documents already circulating inside the company, does it contain an insight a chatbot wouldn't give everyone, and who put in the most effort. Some run submissions through AI detectors. The real test, he says, is "whether your judgment adds something that AI can't produce for anyone else with the same prompt" [1].

Mind the Product makes the same point from the hiring manager's side. If a candidate could pass your case study just by using ChatGPT, the round is measuring chatbot access, not the person [2]. One suggested format is the "audit interview": give the candidate flawed AI-generated work, such as a metrics plan that tracks the wrong thing or a launch plan that skips a team that must approve, and see what they catch. Write the rubric first, or you risk hiring whoever sounds most like you [2].

Gupta also covered offer math. Annualize the sign-on bonus over how long you expect to stay. Value private stock as quoted value × odds of payout ÷

1.15^(years until you can sell) [3]. In his Series C example, the recruiter's $370K first-year figure works out to about $270K a year. That assumes a two-year stay, 40% payout odds and five years to IPO [3]. If you plan to stay 2–3 years, negotiate on sign-on and front-loaded vesting. For 4+ years, negotiate on base and equity [3].

Observed, believed, tested

Ravi Mehta defines data fluency as turning data into insights that change what the team does, then tying them to business outcomes. "Activation fell" is an observation, not an insight [4]. His warning for AI products: if users send more messages because the assistant's answers need correcting, engagement goes up while value goes down [4]. The core skill is knowing whether you've observed something, believe it, or have tested it [4].

Hiten Shah's early-access essay applies the same discipline. Each kind of evidence answers a different question: a signup, a task done with the founder's help, an attempt the user makes alone, and a payment. Even payment doesn't prove the product keeps delivering value [5]. Superhuman's "very disappointed" score went from 22% to 33% just because the team narrowed who it surveyed, before any product changes [5]. On feature requests, Shah asks what the user was trying to do, what stopped them, and how often it happens. He warns that following a few highly engaged users can make the product harder for everyone else [5].

When Customer Success wants the redesign rolled back

On r/ProductManagement, a PM shipped a v2 that changed a legacy workflow. Analytics suggest most users adapted, but Customer Success wants a revert [6]. The strongest reply pointed out that adapting isn't the same as being satisfied, since users may have found workarounds or just stopped complaining. It suggested separating three questions: is v2 better at the task, did the migration add avoidable friction, and is that friction big enough to justify a product change [7]. Another commenter recommended agreeing with CS, in writing, on what would trigger a rollback. Examples: an account up for renewal that cites the change, or a task that measurably takes longer in v2. Anything below that bar gets enablement, not a reopened decision [8].

Building AI features and agents

- **Proven/Better/New.** Gupta, citing someone he calls Mark, sorts features into three groups: proven (copies what already works), better (rebuilt for the new platform) and new (never done before). Because AI is already new, keep the "new" share small. A new use case, built a new way, with AI added on top is "too much new" [9].
- **Snaptrude Agent** takes building design from a site to a construction-ready model and lets designers decide where cuts land when a budget

shrinks. The company reports that more than 2,000 early-access designers ran more than 25,000 tasks. It also built AECBench, its own benchmark, because "an almost correct answer is still wrong" [10].

- **SuperMe** raised a $20M Series A led by Forerunner. It turns operators' work into profiles that help you work through a problem, available in group chats, Claude and ChatGPT [11].

Sources

1. How to Ace PM Interview Presentations and Homework in the AI Era
2. Claude Can Now Pass Coding Interviews. Now What?
3. substack
4. Fluency with Data.
5. X article by @hnshah
6. r/ProductManagement post by u/kylelee
7. r/ProductManagement comment by u/cobramullet
8. r/ProductManagement comment by u/bobiq_ai
9. substack
10. X post by @altafganihar
11. X post by @onecaseman