A KEVIN DUNCAN PLAYBOOK

The Jagged Frontier

A grounded look at what AI actually changes about product management, and the research showing why blind trust in it can make you worse at your job, not better.

THE CORE PROBLEM

Neither the hype nor the panic is telling you the truth

Most writing on AI and product management falls into one of two camps: it will change everything, or it's coming for your job. Neither camp has much interest in the boring, useful answer in between.

The boring answer comes from a rigorous piece of research, not a hot take. In 2023, researchers from Harvard Business School ran a field experiment with 758 consultants at Boston Consulting Group, giving them real, complex tasks with and without access to GPT-4.[1] The finding wasn't that AI made everyone better, or that it made everyone worse. It was stranger and more useful than either: AI capability turned out to be uneven in ways nobody could reliably predict in advance.

The researchers called this the "jagged technological frontier." Some tasks, often ones that looked hard, AI handled brilliantly. Other tasks, sometimes ones that looked almost identical in difficulty, it failed at completely, and failed in a way that sounded confident and looked plausible. The frontier between the two isn't a smooth line. It's jagged, and you often can't tell which side of it you're on until after you've trusted the output.

> **Why product managers specifically should care.** The study was run on consultants, not product managers, but the finding transfers directly. Product work is a mix of exactly the two task types the research describes: well-patterned tasks with lots of precedent, and genuinely novel judgement calls with none. Knowing which is which, before you hand it to a model, is the actual skill this moment demands.

WHAT THE STUDY ACTUALLY FOUND

Inside the frontier, AI is genuinely excellent

On tasks that fell inside AI's capability frontier, the BCG consultants using GPT-4 completed around 12% more tasks, finished roughly 25% faster, and produced work rated as meaningfully higher quality by independent evaluators.[1] The gains were largest for consultants who started below the median, AI raised the floor more than it raised the ceiling.

Then the researchers deliberately included one task chosen to sit outside the model's capability. On that task, consultants using AI were 19 percentage points less likely to reach the correct answer than consultants working without it.[1] The model did not simply fail to help. It made trained, capable professionals worse than they would have been on their own.

> **The reason is "fluent failure."** A wrong answer that reads awkwardly gets checked. A wrong answer that reads confidently and well doesn't. The researchers found consultants developed what they called mis-calibrated trust, over-relying on AI exactly where it was weakest, and under-using it where it was strongest, because fluency, not correctness, was what they were unconsciously grading it on.[1]

The determining skill

The single factor that separated the professionals who benefited most from AI wasn't how much they used it. It was how accurately they could tell, in advance, which side of the frontier a given task sat on. That is not a prompting skill. It's a judgement skill, and it's the one this playbook is actually about.

MAPPING IT ONTO PRODUCT WORK

Two kinds of tasks, not one job

Product management is a mix of well-patterned work and genuinely novel judgement calls, in roughly equal measure. Sorting your own week into the two columns below is more useful than any general debate about whether "AI is coming for product management."

INSIDE THE FRONTIER	OUTSIDE THE FRONTIER
— Drafting a PRD from a clear, agreed brief	— Reading whether a stakeholder's "yes" actually means yes
— Synthesising themes across dozens of interview transcripts	— Betting the roadmap when the evidence genuinely conflicts
— Summarising a meeting into decisions and actions	— Negotiating a trade-off between two teams with a history
— Drafting acceptance criteria from an established pattern	— Noticing a user meant something they didn't say out loud
— A first-pass competitor feature comparison	— Owning the call, and its consequences, when it goes wrong

Notice the pattern: the left column has abundant precedent and a checkable right answer. The right column is context-specific, political, and often has no right answer to check against, only a judgement to defend.

TWO WORKING PRINCIPLES

Working with the frontier

01 Delegate the pattern, keep the judgement

Hand over tasks with abundant precedent and a checkable answer freely. That's where the 25% speed gain and the higher-rated output genuinely show up, and there's no virtue in doing that work slowly by hand.[1] But keep the judgement calls, the ones with no clean precedent, no single right answer, and real consequences if you get them wrong. That's not caution for its own sake. It's where the research says AI is most likely to fail fluently and drag your accuracy down with it.

02 Grade the output on correctness, not fluency

Mis-calibrated trust is the actual risk this research identifies, not job displacement.[1] A confident, well-written, wrong synthesis of your user research is more dangerous than an obviously broken one, because nobody double-checks a document that already sounds right. Build the habit of asking "how would I know if this were wrong?" before you act on anything AI produced for a task you haven't independently verified sits inside the frontier.

PUTTING IT TOGETHER

Auditing a week of PM work

Take an ordinary week and sort it honestly. Most weeks look something like this.

TASK	FRONTIER SIDE	WHY	AI'S ROLE
Draft the sprint PRD	Inside	Clear brief, established format, checkable structure	First draft, then human review
Synthesise 40 support tickets	Inside	Pattern-rich, high volume, low individual stakes	Full delegation, spot-check themes
Decide which of 2 conflicting user signals to bet on	Outside	No precedent, judgement call, real consequences	None. Your call, your accountability
Tell an exec their pet feature tested badly	Outside	Political, relational, context AI cannot see	None. This is the job

WHERE THIS BREAKS DOWN

Common pitfalls

- × **Trusting fluency over correctness.** A confident answer is not the same as a right one. This is precisely the failure mode the research found, and precisely the one people keep repeating.
- × **Refusing to delegate anything.** Treating every task as a judgement call wastes the genuine, measured gains available on well-patterned work. Caution everywhere is its own kind of inefficiency.
- × **Assuming yesterday's frontier is today's.** The boundary moves as models improve. A task outside the frontier six months ago may not be today. Re-check occasionally, don't assume permanently.
- × **Outsourcing accountability along with the task.** You can delegate the drafting. You cannot delegate owning the decision. "The AI suggested it" is not a defensible answer to "why did we ship this."
- × **Letting juniors skip the judgement reps.** AI raises the floor for less experienced performers on patterned tasks, but judgement is built by practising it. Protect room for people to make, and learn from, real calls.
- × **Treating this as a one-off decision.** Which side of the frontier a task sits on isn't fixed. Reassess per task, not per policy.

THE CANVAS

Audit your own week

List five to eight tasks from a typical week. Sort each honestly, not by what you'd like to delegate, but by whether it genuinely has precedent and a checkable answer.

TASK	FRONTIER SIDE	WHY	AI'S ROLE

THE AUTHOR

About Kevin Duncan

Kevin Duncan is a product and digital transformation leader with over a decade of experience delivering product strategy and large-scale change across financial services, government and consumer sectors in the UK.

He currently leads product strategy and transformation for the UK mortgage and property ecosystem as Product Director at Novus Strategy & Consulting.

This playbook is drawn from workshops Kevin has run with product teams on making practical, ungimmicky decisions about where AI actually helps.

Want help building a practical AI operating model for your product team?

kevinduncan.co.uk · email@kevinduncan.co.uk · 07891 163404

References

1. Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz, H., Kellogg, K.C., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K.R. (2023, published 2026) "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality." *Organization Science*, 37(2), pp.403–423. Conducted with 758 consultants at Boston Consulting Group in partnership with Harvard Business School.