PMAIRED™

A Practical Framework for AI-Ready Product Management

White Paper

Author: Babalola Oladeji
Date: January 22, 2025

For Product Managers and product professionals building AI-powered and AI-enabled products.

Executive Summary

Product teams are being asked to "add AI," while simultaneously remaining accountable for the "old" fundamentals: retention, revenue, reliability, and speed to market. This tension has created a new failure mode: "AI features that demo well but don't hold up in prod."

Adoption trends indicate that the rate of adoption continues to rise. McKinsey found that 65% of respondents indicated their organizations were using generative AI on a regular basis by the early part of 2024. IBM found that 42% of companies of enterprise scale had actively deployed AI. In addition, 40% were also looking or experimenting with AI. McKinsey also found that 23% of respondents indicated their organizations were scaling an agentic AI system somewhere across the enterprise by their survey conducted in 2025. However, there are many organizations where the capture of benefits does not keep pace with the adoption of AI because the teams are not yet at a product-grade operating discipline for evaluation, data, and risk.

PMAIRED™, as a practical framework, helps product managers develop a common vocabulary as well as an operating system to work with AI product development. There are seven pillars to this framework: Problem framing, Metrics, AI architecture, Information/data, Risk/reliability/trust, Ethics/explainability, and Delivery/operations.

This white paper will discuss the concept, why it matters, how to apply it throughout the product lifecycle, and finally, how to apply it in 90 days. It will include tables, graphs, mini-caselets such as Klarna's support assistant, as well as an in-depth composite case study, or as I call it, an internal postmortem, as "the truth" often resides here.

Figure 1. Enterprise AI adoption signals (selected surveys/polls).

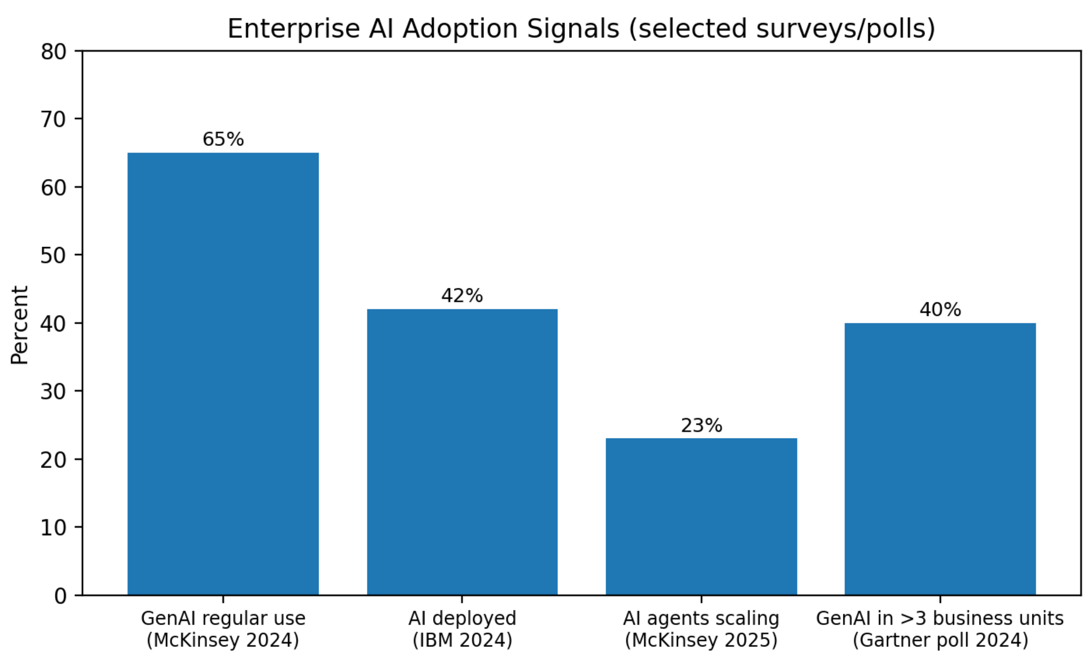


Table of Contents

1. The Product Management Problem AI Creates (and Exposes)

But AI is not just another feature type. AI changes how we decide, build, ship, and operate. The biggest mistake I see again and again is thinking of AI as a ‘module’ you add to your roadmap, rather than a system that changes your product operating model.

But that’s not all: this is where things get uncomfortable – AI doesn’t cause most product issues; it reveals them. Meaning, if your existing processes are subpar with regards to product discovery, data, metrics, or delivery, AI will just amplify all that.

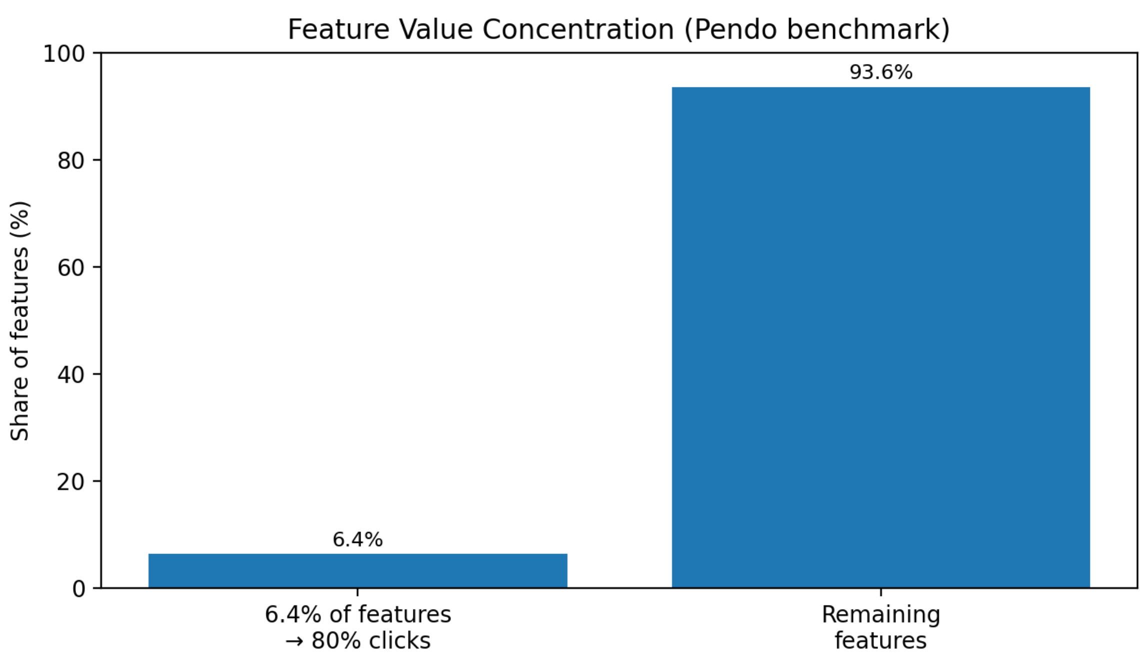
Speeding ahead

The hype curve for AI encourages speed. However, trust with a product isn’t built with demo speed. Trust with a product is built with the way the product behaves.

Feature waste is the baseline problem

Before the emergence of artificial intelligence, many groups had already been creating an excessive number of features that do not attain utility. Pendo developed its Feature Adoption Report, which showed that 80% of features of an average application are infrequently used or unused. Subsequently, Pendo carried out benchmarks that showed an evident distribution of use, whereby many features of an application are dominated by a small number of features. Artificial intelligence will assist in lessening feature waste but will also worsen waste. This is because it will accelerate learning but will worsen judgment.

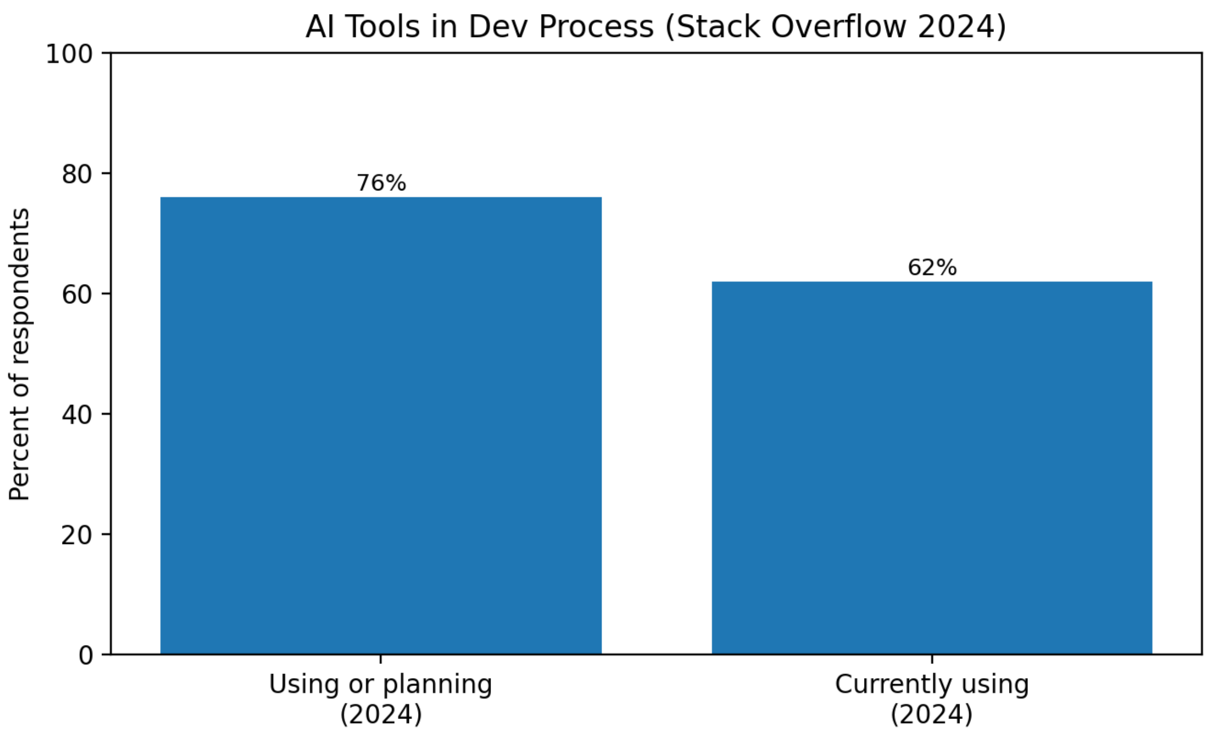
Figure 2. Feature value concentration (Pendo benchmark).



The trust gap is now a product constraint

Even in software development, where AI tooling is mainstream, trust is uneven. The Stack Overflow 2024 Developer Survey reported that 76% of respondents were using or planning to use AI tools in their development process, and 62% were currently using AI tools. Product teams face the same dynamic: usage grows faster than shared evaluation discipline.

Figure 3. AI tools in the development process (Stack Overflow 2024).



2. Defining PMAIRED™

PMAIRED™ is a product management framework for building and running AI-enabled product experiences and AI-enabled product teams. It is designed for the product manager who is tasked with decisions in a context filled with uncertainty about what to build, how to measure success, how to manage risk, and how to ship responsibly.

Moreover, it is cross-functional. PMAIRED™ does not replace engineering design documents and machine learning research; instead, it adds value to these approaches by ensuring product clarity, customer value, and accountability.

2.1 PMAIRED™ pillars at a glance

Letter	Pillar	What it means (PM view)
P	Problem Framing	Define the decision, constraints, and what "good" means; prevent prompt-led drift.
M	Metrics & Measurement	Tie AI outputs to product metrics; separate proxy metrics from outcome metrics.
A	AI Architecture Choices	Choose the right pattern: assistant, copilot, recommender, agent, or autonomous workflow.
I	Information & Data Stewardship	Data quality, governance, consent, privacy, and model-data lineage.
R	Risk, Reliability & Trust	Failure modes, safety boundaries, evaluation, red teaming, monitoring.
E	Ethics & Explainability	Bias, fairness, transparency, user control, responsible deployment.

D	Delivery & Operations	LLMOps/MLOps, rollout, change management, cost controls, incident response.

2.2 What PMAIRED™ is not

• Not a prompt library. Prompting is useful, but it doesn't substitute for product judgment.

• Not an ML engineering framework. It assumes engineers and data scientists have their own practices.

• Not a compliance-only checklist. It balances speed with responsibility and measurable value.

2.3 A simple definition you can use internally

PMAIRED™ is a seven-part discipline that helps product teams choose the right AI pattern, define success, govern data and risk, and operate AI features in production so they create measurable customer and business outcomes.

3. The Business Case for PMAIRED™ (Data, Trends, and Risk)

The investment in AI is increasing, despite some organizations not getting satisfactory ROI on their AI investments. This is the space that PMAIRED™ fills, where AI work is transformed into product work, so that organizations don't get caught up in the activity rather than the actual achievement of results.

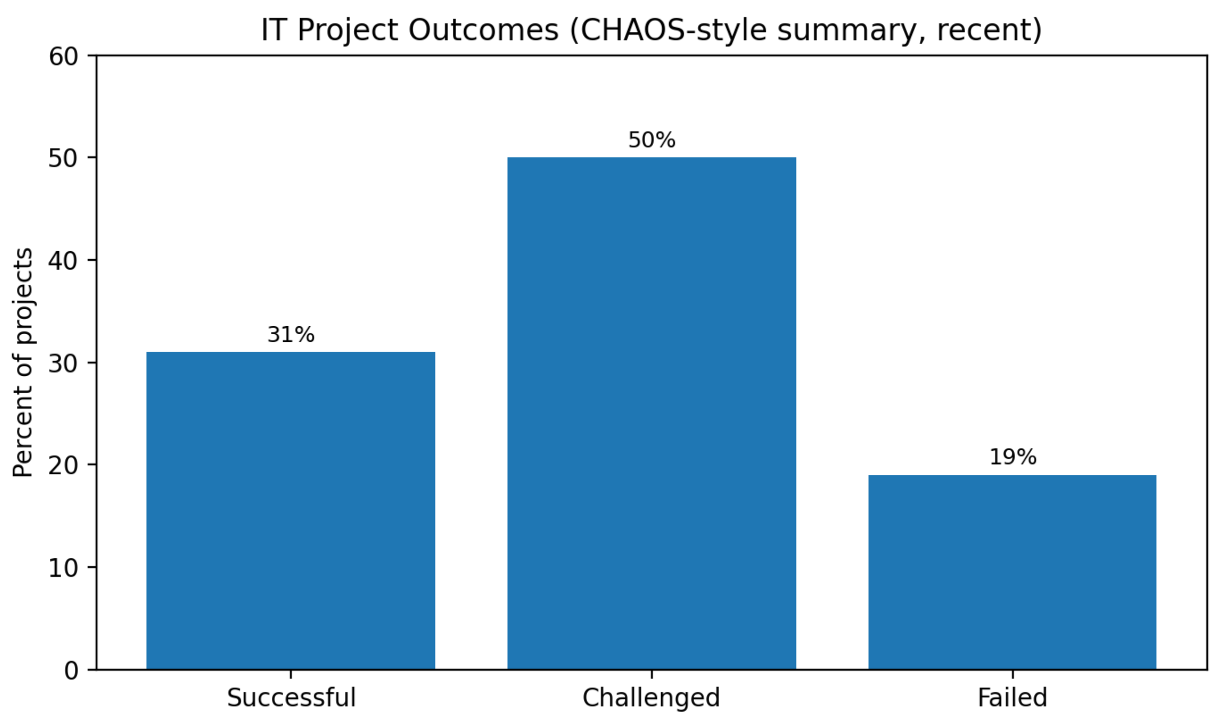
3.1 Adoption is mainstream, scaling the challenge

This can be confirmed through the results of McKinsey's surveys on artificial intelligence, which revealed widespread adoption of generative artificial intelligence. IBM's adoption index also revealed that many big companies are already adopting artificial intelligence, and many are in the stage of experimentation. This can be observed: artificial intelligence adoption is widespread, yet efficient use of it is still scarce.

3.2 Why Projects Fail: Scale, Scope, and Weak Governance

Large AI projects tend to have similarities to large IT projects: they are complex, multifunctional, and have many hidden dependencies. CHAOS analysis still shows many projects face challenges or fail. The lesson for product managers is not to avoid AI projects, but to reduce batch sizes, add gates, and continually measure.

Figure 4. IT project outcomes (CHAOS-style summary).



3.3 Cost is a product variable now

With AI, cost is not just a backend concern. Inference, retrieval, and evaluation costs can scale with usage. That means PMs need cost-to-serve metrics alongside adoption metrics. If you don’t do this, you can accidentally ship a ‘successful’ feature that loses money as it grows.

4. PMAIRED™ Pillars: A Deep Dive

P — Problem Framing

However, for AI, cost is not just a backend problem. There can be cost associated with inference, retrieval, and even evaluation that can scale up along with adoption and hence cost to serve needs to be tracked along with adoption for product managers. If we don't do that, we may end up launching a feature that seems successful but is actually resulting in losses due to increased adoption.

M — Metrics & Measurement

Moreover, the additional failure modes that can be present with AI features include silent failure. Hence, the need for evaluation is not a choice; it has to be a part of the product.

The five dimensions of measurement, defined by PMAIRED™, are behavioral quality, safety, results for users, results for business, and economics, to prevent the common error of measuring usage only.

Table 1. Evaluation dimensions PMs should track.

Evaluation dimension	Example measures
Behavioral quality	Correctness vs ground truth; citation accuracy; refusal quality; hallucination rate
Safety	Toxicity; PII leakage checks; prompt injection resilience; jailbreak resistance
User outcomes	Task completion rate; time-to-value; satisfaction; adoption and retention
Business outcomes	Revenue lift; cost-to-serve reduction; churn impact; risk reduction
Economics	Cost per successful task; latency percentiles; token/compute spend per active user

A — AI Architecture Choices

The term "AI" can refer to several patterns. Some are functionally similar to low-risk assistants, while others are agents that can change the state of the system. In PMAIRED™, the architecture choice was a product decision because it affected the cost and risk.

A working principle is that the smallest pattern that is adequate for the task should be used. Movement through assistant, copilot, recommender, and finally into an agent should only happen when the value and control are demonstrable.

I – Information & Data Stewardship

Data quality, consent, and governance are what define the safe abilities of the product. This is not something that you leave to your lawyer. This is a design input.

Data minimization, data provenance, access control, retention policies, and the distinction between training data and user data during inference are all mandated in PMAIRED™.

R — Risk, Reliability & Trust

Reliability is user experience. If an AI system fails, it feels like a betrayal to the user, not like a bug. This is significant, as it affects user engagement.

PMAIRED™ defines trust using risk levels, guard rails, refusal actions, escalation routes, and monitoring. Incidents are part of the life cycle; they are not unplanned events.

E — Ethics & Explainability

Ethics is not an abstract concern; rather, we are dealing with tangible harm, including bias, unfairness, manipulation, and lack of end-users' agency. Explainability does not imply telling users what's going on inside the model. Explainability means providing users with truthful information about what happened, why (at a level they can understand), and what controls they have. D – Delivery & Operations A "living system," AI is subject to "drift," changes in delivery costs, and changes in user behavior. The absence of operations means "slow-motion failure." A release gate, rollout, and off switch have all been included in the PMAIRED™. The most effective teams also have a post-launch review as a regular activity, like a sprint retrospective.

4.1 PMAIRED™ maturity model (how teams evolve)

Most organizations will move through maturity levels. The goal isn't to jump to Level 5 overnight; it's to know where you are and what the next step should be.

Maturity level	Typical behavior	What works	What breaks
Level 1: Ad hoc	Prompting in private	Individual productivity; no shared evaluation	Risky leakage; inconsistent decisions

Level 2: Pilots	Team prototypes	Basic metrics; limited rollout	Success definitions still weak
Level 3: Productized	AI feature shipped	Instrumentation + monitoring	Incidents handled but not systematized
Level 4: Governed	Multiple AI experiences	Formal evaluation gates + risk tiers	Clear owners; improved trust
Level 5: Operating system	AI across portfolio	Continuous eval, cost governance, model lifecycle mgmt	Durable value capture; scalable velocity

5. Applications Across the Product Lifecycle

PMAIRED™ is designed to be used across discovery, strategy, prioritization, delivery, and post-launch operations. The table below maps common AI-assisted workflows to the primary value and the controls that prevent predictable failure modes.

PM phase	AI-assisted workflow	Primary value	PMAIRED™ controls
Discovery	Interview synthesis and insight clustering	Faster pattern finding; reduces manual fatigue	Link every insight to verbatim quotes + timestamps
Strategy	Market scan and competitor mapping	Broader coverage; faster brief creation	Verify claims against primary sources; citation requirement
Prioritization	Scenario planning and impact narratives	Faster trade-off exploration; better narratives	Avoid false precision; use ranges + assumptions
Delivery	User story drafts and acceptance tests	Less blank-page time; stronger consistency	Peer review + test-first discipline
Quality	Synthetic test generation and edge-case discovery	Finds corners humans forget	Security review; prevent data leakage
Support	Agent assist and draft responses	Lower handle time; consistent tone	Policy filters + escalation triggers
Growth	Personalization hypotheses and copy variants	More experiments; faster learning loops	Holdout tests + guard against manipulation

5.1 A note on ‘AI for prioritization’

However, prioritization carries a risk to product managers (PMs), as it can create a false sense of certainty. The model can produce clear scores and reasoning that look rigorous on the surface, while it might be merely a form of "storytelling." The application of the PMAIRED™ requires the clear articulation of assumptions, the clear naming of confidence, and the comparison to actual historical results as a way to validate the output.

5.2 A note on “AI agents”

Currently, agentic workflows represent the frontier. Agentic workflows are also associated with an increase in product risk, as agents are able to carry out multi-step actions. It is crucial to ensure that, during the deployment of agents, permissions, restricted tools, logging, and human reviews are implemented.

6. Implementation Playbook: PMAIRED™ in 90 Days

This is a practical approach for roll-out for real-world teams with limited bandwidth. The goal is to create a single AI experience that is measurable, safe, and scalable.

**Day 1 to 15: Alignment and Decision Brief

- Chose one workflow that is valuable (support, onboarding, analysis, or internal knowledge retrieval).

- Prepare the PMAIRED™ Decision Brief.

- Establish Baseline Metrics and Target Outcomes.

- Select the AI pattern (start with a smaller scope than final aspirations).

- Determine the sources and constraints on the use of the data (privacy, security, etc.).

Days 16-45: Prototype With Inbuilt Evaluation

- Make a thin slice and instrument all components.

- Create a small ground truth data set.

- Offline evaluation definition and execution.

- Red Team testing for prompt injection and data leakage.

- Determine refusal behavior.

Days 46-75: Controlled Roll

- Internal testing (Dogfooding).

- Roll-out to a limited cohort.

- Monitor quality, safety, and cost-effectiveness.

- Tackle the top three failure modes before large-scale scaling occurs.

- Document lessons learned as a reusable playbook.

Days 76–90: Scale decision and operating cadence

Scale

- Compare the result with the baseline. - Decide to scale, pivot, or discontinue. - Establish Ongoing Monitoring Dashboards. - Implement monthly review of risks and costs. - Identify an owner for ongoing model/experience health.

7. Operating Model: Governance, Cadence, and Ownership

In many companies, it's been observed that positional ambiguities are common after the implementation of an AI feature. PMAIRED™ resolves this problem through an established cadence and an ownership model.

7.1 Cadence (simple and non-negotiable)

• Weekly: Quality and Incidents – Top failures, root causes, fixes.

• Biweekly: Metric review (user outcomes, business outcomes, economics).

• Monthly: Governance review (reassessing tier status for identified risks, policy changes, and changes in

• Quarterly: Strategy review (identifying areas where AI has sustained competitive advantage vs. mere novelty).

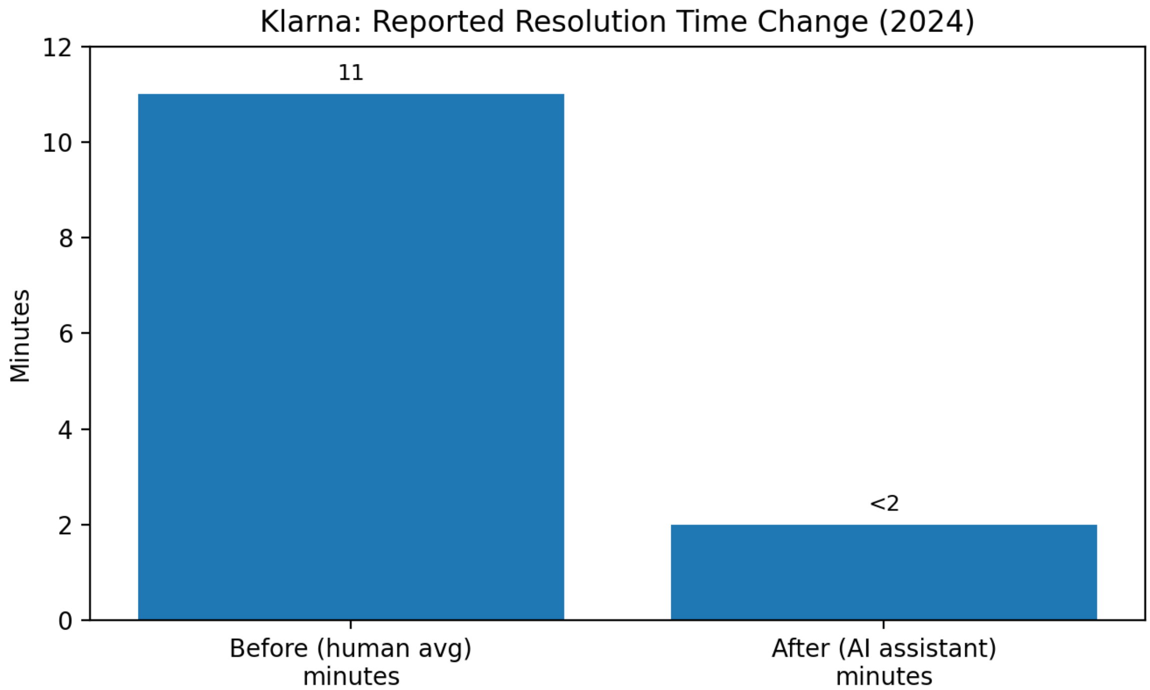
7.2 Ownership map (RAC

The product is responsible for outcomes and risk. Engineering takes charge of implementation and performance. Data and ML are in charge of evaluation and monitoring. Lastly, we have security and privacy, which are responsible for threat models and policy compliance. We have design, which takes charge of transparency and user controls.

8. Case Studies 8.1 Real-World Klarna (Customer service):

Klarna announced that, in February 2024, two-thirds of customer service conversations, equivalent to 2.3 million conversations and the work of 700 full-time staff, had been handled by Klarna’s AI assistant in the first month of deployment. Klarna also announced that the time taken to solve customers’ issues had fallen to under 2 minutes, compared to 11 minutes prior to this, and that 25% of recurring questions had been reduced.

Figure 5. Klarna: reported resolution time change (2024).



GitHub Copilot (developer productivity): GitHub's research reported that developers using Copilot completed a coding task 55% faster than those who didn't in a controlled experiment. This matters for product leaders because it shows the shape of AI value: measurable time savings on well-defined tasks, not magic on ambiguous ones.

8.2 Composite Case Study: NorthstarHR Support Copilot

Method Note: This composite case integrates most common trends that are often seen in mid-market SaaS companies. It is designed to be realistic but avoid revealing confidential information that might be associated with any single company.

Context

NorthstarHR is a B2B HR operations platform used in mid-sized businesses. The cost of support services is rising, and customer satisfaction levels are decreasing. The team has chosen to build a support copilot for support agents first, followed by a customer-facing copilot.

P: Problem framing

The Product Manager will develop a one-page decision brief. The goals are: reduce median time-to-resolution by 20%, deflect 15% of top repetitive intents' tickets, and prevent an increase in compliance incidents. The non-goals will be clearly defined: no legal guidance will be provided, no account changes will be autonomous, and no use of sensitive fields will be allowed in prompts.

M: Metrics and evaluation

The accuracy, citation accuracy, and escalation accuracy are evaluated using a 300-ticket dataset. Product metrics, such as deflection, handle time, and CSAT, are tracked, along with cost per resolved ticket.

A: Architecture choice

The selection of RAG-based grounding over internal knowledge bases is made, with strict citations enforced. Customer deployment is postponed until copilot stability is demonstrated.

I: Data Stewardship Sensitive fields are redacted, retention policies are followed, and minimal logging and access controls are performed for data evaluations.

R: Risk, reliability, trust A refusal mechanism is enabled for missing or inconsistent sources. Escalation workflow maintains human accountability. Incident review happens as part of a weekly routine.

E: Ethics and explainability A user experience pattern for "show sources" has been included. This has a concise and honest description. In addition, users have the ability to give feedback on incorrect answers.

D: Delivery and operations Rollout happen in phases: internal dogfooding -> 10% cohort -> full agent rollout -> customer assistant for top intents.

Outcome after 8 weeks: 18% reduction in median time-to-resolution and 12% deflection for top intents, zero material compliance issues.

9. Templates, Checklists, and Metrics Library

9.1 PMAIRED™ Decision Brief (one-page template)

1) User and job to be done

2) Problem statement (what pain, for whom, when)

3) AI Pattern - AI can be

4) Baseline and target metrics (user outcomes and business outcomes)

5) Non-goals (explicit boundaries

6) Data sources and constraints (privacy, retention, access)

7) Risk Tier and Controls (Human-In-The-Loop, Refusal,

8) Evaluation plan (ground truth and offline tests)

9) Rollout plan (Shadow mode, Cohort rollouts, etc.)

10) Owner and Operating Cadence

9.2 Release Gating Checklist

• A ground truth dataset is available. The dataset is versioned.

• Offline evaluation is satisfactory.

• Prompt injection and data leakage tests have been performed.

• UX transparency patterns implemented (sources, refusal, confidence).

• Monitoring dashboard is live prior to rollout.

• The off switch is available and has already been tested.

• Incident response owner has been assigned

9.3 Risk classification quick table

Risk level	Examples	Minimum controls
Low	Brainstorming, internal summaries, drafts	Disclosure + basic monitoring
Medium	Customer support drafts, recommendations, analytics narratives	Source grounding + human review option + audit logs
High	Financial decisions, access changes, eligibility decisions	Human-in-the-loop required + formal eval + approvals + ongoing audits

10. Conclusion

PMAIRED™ has been built for the world product managers face today: AI adoption is the new normal, trust is hard won, and AI value is unevenly distributed. It will not be the teams that can deliver the most compelling demos that will succeed, but the teams that can deliver AI experiences that are predictable, measurable, and governable.

The implementation of PMAIRED™ will prove to be enlightening, revealing to you that there is a significant, if limited, outcome: your team will start to ask more effective questions. What decision are we supporting? How do we validate our success? What do we do if it doesn't work? What is the cost per success? While these questions may be considered simplistic, they actually define the boundary between novelty and advantage.

References

McKinsey & Company. (May 30, 2024). The state of AI in early 2024. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

McKinsey & Company. (Nov 5, 2025). The State of AI: Global Survey 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

IBM. (Jan 10, 2024). Data Suggests Growth in Enterprise Adoption of AI is Due to Widespread Deployment by Early Adopters. https://newsroom.ibm.com/2024-01-10-Data-Suggests-Growth-in-Enterprise-Adoption-of-AI-is-Due-to-Widespread-Deployment-by-Early-Adopters

IBM & Morning Consult. (Dec 2023 / published Jan 2024). IBM Global AI Adoption Index Report (PDF). https://filecache.mediaroom.com/mr5mr_ibmspgi/179414/download/IBM%20Global%20AI%20Adoption%20Index%20Report%20Dec.%202023.pdf

Gartner. (Early 2024). What Generative AI Means for Business (poll result summary). https://www.gartner.com/en/insights/generative-ai-for-business

Pendo. (2019). The 2019 Feature Adoption Report. https://www.pendo.io/resources/the-2019-feature-adoption-report/

Pendo. (Jul 1, 2024). Feature Adoption Benchmarking. https://www.pendo.io/pendo-blog/feature-adoption-benchmarking/

Pendo. (Nov 25, 2019). Tactical vs. Strategic: Where Product Managers Really Spent Their Time in 2019. https://www.pendo.io/pendo-blog/tactical-vs-strategic-where-product-managers-really-spent-their-time-in-2019/

Stack Overflow. (2024). Developer Survey 2024: AI tools in the development process. https://survey.stackoverflow.co/2024/ai

Klarna. (Feb 27, 2024). Klarna AI assistant handles two-thirds of customer service chats in its first month. https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/

Reuters. (Feb 28, 2024). Teleperformance shares plunge on AI disruption concerns (Klarna figures cited). https://www.reuters.com/technology/teleperformance-shares-plunge-ai-disruption-concerns-2024-02-28/

GitHub. (Sep 7, 2022). Research: quantifying GitHub Copilot’s impact on developer productivity and happiness. https://github.blog/news-insights/research/research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/

OpenCommons. (May 14, 2025). CHAOS Report on IT Project Outcomes (summary). https://opencommons.org/CHAOS_Report_on_IT_Project_Outcomes

Appendix A: AI Patterns for Product Managers (Assistant → Agent)

Assistant

Definition: Generates or transforms content for a human user (write, summarize, ideate).

Best for: speed on low-risk tasks.

Common failure: confident misinformation.

PMAIRED™ guardrails: source grounding when facts matter; disclosure; human review on external-facing output.

Copilot

Definition: Suggests actions inside an existing workflow (e.g., support agent drafts, PM story drafts).

Best for: reducing cycle time in repeatable processes.

Common failure: brittle suggestions that don't match context.

PMAIRED™ guardrails: constrained context; templates; mandatory edit step; feedback loop.

Recommender / Ranker

Definition: Predicts or ranks options (next best action, prioritization signals).

Best for: surfacing patterns in messy data.

Common failure: proxy metrics become the goal.

PMAIRED™ guardrails: bias tests; calibration; compare to human baselines; monitor drift.

Agent

Definition: Plans and executes multi-step tasks, often via tools (search, write, call APIs).

Best for: complex workflows where autonomy creates real leverage.

Common failure: unsafe tool use, action without permission, or 'runaway' costs.

PMAIRED™ guardrails: least-privilege tools; approvals; step logging; rate limits; hard stop conditions.

Appendix B: Common Failure Modes (and What PMAIRED™ Does About Them)

Failure mode	PMAIRED™ mitigation
Hallucinated facts in customer-facing content	Require source grounding, refuse if sources are not available, and add human review for high-risk messages.
Data leakage in prompts/logs	Minimize and Redact, Enforce Strict Retention, Access Controls, and Security Review for prompt injection/exfiltration.
False precision in prioritization	Use ranges, not single scores; require assumptions; compare against historical outcomes; label confidence.
Silent quality drift after launch	Model performance metrics should be monitored, and re-evaluation should be scheduled after data and policy changes; regressions should be handled as incidents.
Cost blowouts as adoption grows	Track cost per successful task, implement caching, use rate limits, and define outcome-based budgets.
User distrust and abandonment	Employ transparency UX patterns, e.g., display sources, display uncertainty, and offer controls.
Governance paralysis	Use risk tiers; keep controls proportional; ship in controlled cohorts with measurable gates.

Appendix C: Security and Privacy Checklist (PM-Friendly)

The program managers do not necessarily need to be security engineers but do need to know what is needed before launch. This checklist is provided for your use.

- Threat modeling complete: prompt injection, data exfiltration, unauthorized tool usage.

- Handling rules for PII/PHI/PCI documented and tested.

- Redaction and minimization of prompts and logs.

- Access controls that are applied on the training and/or evaluation data.

- Retention policy defined (duration of prompt/log storage, authorized access).

- Incident response plan defined (owners, escalation, communication to users). - Abuse monitoring (spam, jailbreaking, adversarial prompting).

Appendix D: Metrics Library (Examples)

Use these as a starting menu. Pick a small set per feature; avoid metric sprawl.

Quality

• Answer correctness rate

• Citation accuracy

• Refusal correctness

• Edge-case pass rate

Safety

• PII leakage rate

• Jailbreak success rate (red-team)

• Policy violation rate

• Security incident count

User outcomes

• Task completion rate

• Time-to-value

• User satisfaction (CSAT/NPS)

• Adoption and retention

Business outcomes

• Support cost reduction

• Revenue lift

• Churn reduction

• Risk reduction indicators

Economics

• Cost per successful task

• Latency p95

• Tokens per session

• Cache hit rate

Appendix E: Mini-Caselet Addendum (Customer Support Agents)

Intercom's Fin AI Agent is an example of an AI support product where the price is charged based on the conversations resolved instead of the conversations used. This type of payment structure is consistent with the concept promoted by PMAIRED™, where payment is made when the desired outcome for the customer is reached. According to the publicly available help documentation for Intercom's product, the price for the Fin AI Agent is charged at $0.99 per resolution for customers in the United States.

The implications for product management: This framework forces the product management definition of success into operation. If the AI does not solve the problem, it is not a success, and it should not quietly inflate the cost

Whitepaper Written by

Babalola Oladeji.

Product Manager