

Building a Generative AI Shopping Inspiration Experience

How I led the development of an LLM-powered stylistic inspiration product that pairs multimodal retrieval with generative content — and the product decisions that made it shippable.

Context

At a large home-goods e-commerce company, customers browsing for furniture and décor often lack the vocabulary to describe what they want. They know a style when they see it, but keyword search and traditional recommendations can't bridge the gap between “I want my living room to feel like this” and a concrete set of products. As Senior PM for Recommendations, I owned the question: could generative AI close that inspiration gap?

The Problem

Traditional recommendation systems retrieve items similar to what a customer has already engaged with. They cannot compose a coherent aesthetic point of view, explain why items belong together, or generate the inspirational content — imagery and styled sets — that helps a customer commit to a vision. Editorial teams could do this manually, but only for a handful of styles, refreshed a few times per year. We needed inspiration at personalization scale.

Approach & Key Decisions

- **Retrieval-first architecture.** Rather than asking an LLM to “know” the catalog (a hallucination risk with millions of SKUs), we grounded generation in multimodal, embedding-based retrieval: visual and semantic embeddings surface stylistically coherent candidate sets, and the LLM layer composes, names, and narrates the inspiration set on top of real, in-stock products.
- **Generative where it adds taste, deterministic where it must be true.** Product facts — price, availability, dimensions — never pass through the generative layer. The LLM owns tone, styling logic, and imagery direction; the retrieval and catalog systems own truth. This split was the single most important design decision for both quality and trust.
- **Evaluation before scale.** There was no off-the-shelf metric for “is this inspiring?” We built a layered eval approach: automated checks for coherence and brand safety, human review rubrics for stylistic quality, and behavioral metrics (engagement, add-to-cart from inspiration surfaces) as the ultimate arbiter.
- **Cost and latency as product constraints.** Generative imagery and LLM calls are expensive. We designed for pre-computation and caching of inspiration sets with online personalization of which sets a customer sees — keeping serving costs and latency compatible with high-traffic surfaces.

Impact

Customer experience	Personalized, stylistically coherent inspiration sets with generated content — a net-new discovery surface, not an incremental widget.
Business	Meaningful incremental engagement and revenue on inspiration-driven journeys (directional; figures confidential).
Organization	Established reusable patterns — grounding, eval rubrics, cost envelopes — that de-risked subsequent GenAI initiatives across the org.

What I'd Tell Other PMs

“The hard part of GenAI products isn't the demo — it's deciding what the model is allowed to be wrong about.”

Scope the generative layer to the parts of the experience where taste matters and errors are cheap. Ground everything else. And invest in evaluation from day one: if you can't articulate what “good” looks like in a rubric, you can't ship, iterate, or defend the experience.