Mihir Parmar

Principal Product Manager | AI Product Builder: Agents, Evals, Shipped Product

+1 (908) 745-9085 | parmarmihir@gmail.com | San Francisco Bay Area

mihirparmar.ai | linkedin.com/in/mihir-r-parmar | github.com/mihirparmar-ai

SUMMARY, SKILLS & TOOLKIT

Product leader who builds the thing. In my own business I shipped and operate a client-facing AI product alone with Claude Code as the pair: a **React/TypeScript** site and itinerary pages on **Supabase and Vercel**, a flight-search **MCP server**, three kinds of agents (itinerary building, destination research, client questions) and the **versioned eval rules** that check their work, all live for paying clients (**$700K+ in bookings in 2026 to date**). At Zillow I owned multimodal LLM document understanding from demo to **1M+ documents a year** and the eval practice that gated it. Ex-Goldman engineer, ex-Deloitte consultant. I prototype before I spec, write SQL to answer my own questions, read the model outputs, and judge myself on what is live in production. Returning to a full-time product seat; En Route Luxe continues as a bounded advisory run on the agent stack I built.

- **Build:** Claude Code daily (custom skills, hooks, MCP servers) • TypeScript/React, Node, Python, SQL • Supabase • Vercel • GitHub (own repos, 110+ PRs, Actions) • Lovable scaffolds • ElevenLabs, Runway, HeyGen, Wispr Flow for content
- **AI systems:** Agents that take actions for users (tool use, MCP, context engineering, memory) • Evals (golden sets, calibration, precision/recall, ship/no-ship, LLM-as-judge vs human raters vs code checks) • Prompt and rules versioning • Failure-mode catalogues • Guardrails and act/draft/escalate policies • Retrieval over a structured corpus • Model selection, build vs buy, cost per task
- **Product:** 0 to 1 with real customers • Customer discovery & field research • Experimentation & instrumentation • Taste: rejecting output that works but should not ship • Platform & API products • Executive storytelling
- **Domains:** Real estate transactions, rentals, mortgage, title and closing • FinTech and trading • Contracts and lease documents • Luxury travel & hospitality

PROFESSIONAL EXPERIENCE

En Route Luxe - Founder / Product Lead (solo, AI-operated luxury travel advisory) | SF Bay Area | 2025 - Present:

The production environment where I design, build, evaluate and operate an agentic system for paying clients, alone: **20+ clients a year, 50+ trips, 30+ countries, $700K+ in bookings in 2026 to date**, grown entirely on repeat and referral.

- **Shipped the product surface myself, no engineering headcount:** the client site and itinerary pages (**React/TypeScript, Supabase, Vercel**; 110+ pull requests, every one reviewed and merged by me), a flight-search **MCP server** any agent in the stack can call, and custom **Claude Code skills** for recurring workflows. All live and client-facing today.
- **Built and operate the connected advisor workspace that runs every trip:** one trip record carries the client brief, approved destination, sourced research, versioned itinerary, supplier quotes, approved proposal, financial obligations and due dates; a coordinator agent routes bounded tasks to a planning agent, a document-and-sourcing agent and a read-only reviewer; a changed preference or supplier term reopens only the decisions it affects.
- **Supplier document intelligence, the Zillow discipline applied to DMC quotes:** ingests PDFs, email and spreadsheets, extracts the commercial record (dates, nights, rooms, rate basis, inclusions, taxes, currency, validity, payment and cancellation terms) with every material field linked to its source, normalizes scope before comparing, flags omissions and revisions; code checks totals, dates and required fields, and a parsed quote, an approved proposal and a confirmed booking stay distinct states.
- **CRM and financial follow-through on the same record:** pipeline stage and next action per client, invoices prepared from approved terms, receipts matched to obligations with partial, duplicate and unmatched records routed to review, balances and deadlines derived from approved terms.
- **Three kinds of agents in use for paying clients:** itinerary-building agents that compile a family's profile into constraints and draft the day-by-day itinerary and client page; destination-research agents over a **190-destination corpus**; and client-query agents that answer questions from the trip record (when a client on the road asked whether to book a **~$7K** set of flights, the agent found the same routing for **~$2,000 less**). Most of this year's trips started as an agent draft I reviewed before it reached the client.

- **Evals, guardrails and human-in-the-loop by design:** a golden set of trips with expected findings, a checker scored on recall and precision, and seven failure classes from production turned into **versioned, dated rules** (closures on the day scheduled, timed ticket windows, transfer buffers, stale venues, verified versus estimated pricing), plus quote and finance acceptance cases. Every action class is **act alone, draft, or escalate**: the model drafts, Gmail permission is create-draft only, invoicing and money movement need my approval, every site change goes through a pull request I merge; judgment calls stay human. Public builder with golden set and eval command: **github.com/mihirparmar-ai/agent-skills/itinerary-builder**.

Zillow Group - Principal Product Manager, AI Document Intelligence Platform | SF Bay Area | 2019 - 2026:

Built Zillow's document intelligence platform 0 to 1: one multi-tenant foundation for every document, business, and internal team; extraction from OCR to multimodal LLMs; an AI-first product for brokers and agents on top.

- **Shipped multimodal LLM document understanding to production and chose the gate that made it trustworthy:** average accuracy was the wrong target, because a contract that is 92% right is unusable if nobody can tell you which 8% is wrong, so we gated on **per-field confidence** with human review on low-confidence fields. **~85%** to **98%+**, **~100% on reviewed compliance-critical fields**, **~$2M a year** of manual review absorbed and the outsourced review workforce cut by more than half, across **1M+ documents a year** and **20+ teams**.
- **Built and owned the eval practice that gated every rollout:** golden datasets drawn from **millions of human-reviewed documents across 100+ document types**, thresholds calibrated per field by confidence bucket and set by the cost of a wrong answer, human labeling on critical fields and code validation on structural ones, offline regression gates on every model, prompt and vendor change, and production drift monitoring. The gates had teeth: a vendor update that raised average accuracy but hallucinated values into blank fields was held until the field-presence check shipped.
- **Zillow Rentals (flagship proof at scale):** shipped digital leases nationally in ~12 months; instrumented and analyzed the create, send and sign funnel in SQL and iterated with experiments. Funnel analytics showed the drop-off was at lease creation, not signing, so the work went into creation, led by **Zillow Leases**, a customized digital lease generated per property, tenant and term: **doubled landlord activation**, **market share 18% to 30%**.
- **Built the multi-tenant Document Platform first, then deployed it tenant by tenant:** replaced DocuSign and dotloop with in-house e-sign, storage, compliance and fillable forms and rolled the platform into **every Zillow business and internal team that touched a document** (Rentals, Home Loans, Zillow Offers, Zillow Closing Services, Premier Agent, external real estate brokerages, plus marketing and sales, legal, recruiting, vendor contracts; **20+ teams**), each a deployment with its own **VP or C-level sponsor**, baseline, success metric, and adoption playbook. Recurring asks became platform capabilities (fillable forms from MLS templates, PII redaction across every document type) or were declined: **1M+ docs/year** at **95%+ platform accuracy**, **~$2M/yr** e-sign licensing retired.
- **Owned model economics and the vendor decision: ~$7M annual budget**; build-vs-buy across 3 vendors + an internal PoC, every candidate scored per field on the same golden sets; chose the vendor with calibrated field-level confidence and an eval harness we could point at our own data, integrated in ~4 weeks vs a ~6-month build, avoided **~$2M/yr** of platform engineering, and stayed model-agnostic. Own the evals, rent the pipeline.

Deutsche Bank, Credit Suisse, UBS | Product Management (banker and prime brokerage platforms) | New York & Singapore | 2013 - 2019:

- Led a **25-person global team** (engineering, design, data) that shipped **iOS and web workflow tools and analytics** for investment bankers across the Americas, EMEA and APAC, raising productivity **~30%** where nothing ships without compliance, legal and risk sign-off.
- Won the **$5M** prime brokerage re-engineering case with a Managing Director sponsor: **10+ hedge fund clients** onboarded, **$50M annual revenue**, **$500K/yr** operating cost cut through automation.

Deloitte | Management Consultant | New York | 2011 - 2013:

- **100+ metric vendor scorecard** that settled **buy-versus-build** and roadmap for a hedge fund's portfolio platform.

Goldman Sachs | Software Engineer | New York | 2007 - 2011:

- Engineer (Java, Perl, SQL) on a global portfolio and trading platform (**$100B+ AUM, 10K+ daily transactions**); built trade-booking automation that cut manual errors and **saved $600K+ a year**.

Education: M.S., Computer Science, New Jersey Institute of Technology (2007) • B.E., Electronics and Telecommunication, Vishwakarma Institute of Technology, University of Pune (2005)