

Abhishek Kolge

Senior AI Product Engineer · Remote, any timezone

abhishekkolge96@gmail.com · +91 97693 13584 · abhishekkolge.dev · github.com/AbhishekKolge · linkedin.com/in/abhishek-kolge

I build AI products end to end, most recently in healthcare, real estate and allergy-safe cooking. I don't let the model make the final call on safety or money, and I test AI features against known answers before they ship.

EXPERIENCE

Senior AI Product Engineer, Contract · healthcare, proptech, consumer iOS Feb 2023 to present

- Rebuilt a healthcare data analyst in 9 weeks with one other engineer, for researchers who never wrote SQL.
- An orchestrator plans each question, then code fans it out to parallel SQL sub-agents, one per dataset.
- Wrote a SQL guard that can refuse a query 24 ways, then closed 7 bypasses I found by probing it myself.
- Scored answers on a golden set that re-runs its own ground truth; it passed 10/10 four runs in a row.
- Built a real-estate deal platform from an empty repo: 31 models, 123 endpoints and 26 durable functions.
- Cut underwriting from days of retyping a rent roll to minutes.
- Constrained extraction to a fixed enum schema, and plain code recomputes each number before anyone sees it.
- Reverted an agent-loop migration after two days: the loop only added failure modes to a fixed sequence.
- Designed three-tier RBAC over 8 roles, where a cross-tenant read returns 404 so nobody can enumerate tenants.
- Wrote 161 Terraform resources by hand without off-the-shelf modules, and one tag push ships every service.
- Deployed to AWS through GitHub OIDC pinned to audience and subject, so no pipeline holds a static cloud key.
- Denied task roles the 13 IAM calls that escalate a breach, wrote WAF into staging and prod, and a DR runbook.
- Rebuilt generation as a resumable polled job after real runs of 30 to 190s hit a 60s load-balancer cut-off.
- Recommended holding a release after finding 9 food-safety defects by testing against the live API.
- Owned the API, AI generation and AWS platform of a consumer iOS app, with 2,363 tests green at release QA.

Sr. Software Developer, DigitalSalt Feb 2022 to Feb 2023

- Took Hedged from first commit to launch in 5 months with four engineers, on web, iOS and Android.
- Handed the app to the client's team, who have since taken it past 100k installs at 4.5 stars.
- Built a RAG assistant over ~8k docs that closed ~40% of support queries, scoped to each asker's permissions.
- Cut latency 25% across Postgres, Redis and RabbitMQ on services handling ~1.5M requests a day.
- Built the monorepo behind 8 apps and the UI library 12 engineers use, and cut build times 40%.
- Shipped zero-downtime releases on ECS and Lambda.
- Mentored 3 engineers and sat in on 10+ hiring interviews.

Frontend Developer, AppOctet Aug 2021 to Feb 2022

- Built client-facing React and Next.js apps in my first job, delivered on fixed dates.
- Learned to write a screen as a state machine: loading, empty, error, stale, and responses the docs did not promise.

SELECTED WORK

Mosaic AI	Answers plain-English questions about healthcare data with guarded SQL.	case study
Perpetuity	Real-estate deal platform where Claude reads documents and people approve numbers.	case study
The Food Maestro	Writes recipes around your allergies on iOS. I owned the API, AI and platform.	case study
Hedged	Trading guidance for first-time investors on web, iOS and Android, 100k+ installs.	case study
ZenAdmin	Automates IT and HR for distributed teams across people, software and devices.	zenadmin.ai
Mahindra Lifespaces	Homebuyer app for a listed developer, with documents, payments and service requests.	Play Store

SKILLS

LLMs	Claude on Bedrock, GPT-5, multi-agent orchestration, tool calling, LangChain, LangGraph, Mastra
Retrieval	RAG, hybrid full-text + pgvector (RRF), ONNX embeddings, text-to-SQL, structured outputs
Guardrails	Golden-set evals, deterministic gates, SQL AST guards, human approval, token ceilings, Langfuse
Product	TypeScript, React 19, Next.js, React Native, Python, FastAPI, Node.js, Postgres, Snowflake
Platform	Temporal, Inngest, AWS, Terraform, GitHub Actions OIDC, IAM boundaries, tenant isolation
Lead	Agentic coding behind written plans and review gates (abhishekkolge.dev/ai-workflow), mentoring