

SHALIN BHATT

Engineering Leader & AI Systems Architect

shalin.dev@proton.me | 415-490-7852 | San Francisco Bay Area
shalinbhatt.dev | github.com/shalin-dev | linkedin.com/in/shalinkb

PROFESSIONAL SUMMARY

Engineering leader with 12 years building and leading teams in regulated enterprise environments — healthcare platforms serving 50K+ daily users, HIPAA-compliant, 99.95% uptime.

Built production RAG pipelines with multi-model routing (OpenAI, Anthropic, Groq) and hallucination mitigation through source-grounded retrieval. Led AI-assisted coding pipelines and owned technical RFCs that set the quality bar across a 12-person team. Build autonomous multi-agent systems, LLM evaluation frameworks, and AI observability infrastructure.

Managed teams of 8-15 engineers with 90% retention. Hired 10+ engineers. Served as interim Engineering Manager during leadership transitions, reporting directly to VP.

Seeking Head of AI Engineering, Engineering Manager, or Staff-level roles where autonomous systems or AI infrastructure is central to the product.

KEY ACHIEVEMENTS

15

Engineers Led & Managed

10

Engineers Hired

90%

Team Retention (vs 60-70% avg)

5+

Promoted to Senior Roles

50K+

Daily Users Served

40-60%

Faster Dev via AI Tooling

99.95%

Platform Uptime

500+

Enterprise Installations

CORE SKILLS

Autonomous Agent Systems: Multi-Agent Architecture, Agent Orchestration, Autonomous Decision Systems, Concurrent Execution & State Management, Self-Healing & Failure Recovery, Tool Use & Function Calling

Production LLM Engineering: RAG, LLM Evaluation Frameworks, LLM-as-Judge Patterns, Multi-Model Routing & Fallback, Prompt Engineering, Context Window Management

AI Reliability & Observability: Agent Workflow Tracing, Production AI Monitoring, Hallucination Mitigation & Guardrails, Custom Evaluation Pipelines, Latency Optimization, Model Benchmarking

AI Infrastructure: Vector Databases & Semantic Search, Memory-Augmented Reasoning, Knowledge Persistence, Containerized AI Deployment, API Design for LLM Systems

Engineering Leadership: Team Management (8-15 Direct Reports), Performance Reviews, Technical Hiring (10+), Roadmap Planning, Cross-Functional Leadership, Mentorship

Technical Stack: Python, LangGraph, LangChain, OpenAI, Anthropic, Groq, Ollama, ChromaDB, FastAPI, Docker, Kubernetes, PostgreSQL, Redis

PROFESSIONAL EXPERIENCE

Engineering Leader | AI Systems Architect | STERIS CORP

Richmond, CA | Dec 2021 – Present

AI-Assisted Engineering & RAG

- Built production RAG pipelines for internal knowledge retrieval and document search — multi-model routing (OpenAI, Anthropic, Groq) with hallucination mitigation through source-grounded retrieval and automated response verification
- Led AI-assisted coding pipelines — automated code review, testing, and documentation generation integrated into daily development workflows
- Reduced AI development cycle time by 40-60% through AI-first workflows and agentic coding tools adopted across a 12-person team

Engineering Leadership

- Managed 8-15 direct reports with 90% retention vs 60-70% industry average — conducted performance reviews, ran 1:1s, and mentored 5+ engineers into senior roles
- Served as interim Engineering Manager during two leadership transitions — reported directly to VP on roadmap, risks, and strategic decisions
- Owned technical RFCs and architecture decisions across multiple product initiatives — set engineering quality standards, drove technical clarity, balanced technical debt reduction with new feature delivery
- Led technical hiring: designed interview process, evaluated candidates, and made 10+ hiring decisions

Platform Engineering

- Led platform engineering serving 50K+ daily users across 500+ enterprise installations with 99.95% uptime in HIPAA-regulated environment
- Designed CI/CD pipelines with Docker and Kubernetes — 70% faster deployments with zero-downtime releases
- Delivered 15+ major feature releases including internationalization for 4 markets — managed cross-functional initiatives with Product, Design, Finance, and Operations

Tech: Python, LangGraph, OpenAI, Anthropic, Groq, Docker, Kubernetes, PostgreSQL, Redis

Senior Software Engineer | STERIS CORP

Richmond, CA | Oct 2017 – Nov 2021

- Owned backend platform architecture supporting multiple product lines — full lifecycle from technical design through production deployment
- Led Scrum process across multiple teams — sprint planning, retrospectives, and technical roadmap ownership driving 15+ feature releases
- Hired and mentored engineers — recruited talent, ran technical interviews, and grew junior developers into senior roles
- Built CI/CD pipelines and testing frameworks for containerized Linux environments — improved deployment reliability and reduced manual overhead

Tech: Ruby on Rails, PostgreSQL, JavaScript, Docker, Kubernetes, Jenkins, Linux

SELECTED AI PROJECTS

Local AI Application Suite (github.com/shalin-dev/local-ai-suite)

Privacy-focused 4-app AI platform running 100% locally: Personal AI Assistant, Code Documentation Generator, Local RAG System, and AI Image Classifier. Fully containerized with Docker.

Tech: React, TypeScript, FastAPI, Ollama, ChromaDB, Docker

LLM Eval Harness (github.com/shalin-dev/llm-eval-harness)

LLM evaluation framework with LLM-as-judge pattern, consistency testing, and RAG evaluation. Built 6 automated evaluators with multi-provider runners.

Tech: Python, Sentence Transformers, PyTorch, Click, Ollama, OpenAI API

AI Observe (github.com/shalin-dev/ai-observe)

AI observability layer for production LLM applications. Decorator-based agent workflow tracing, web dashboard with timeline visualization, and framework integrations for Ollama, CrewAI, and LangChain.

Tech: Python, FastAPI, SQLAlchemy, Jinja2, SQLite, PostgreSQL

EDUCATION & CERTIFICATIONS

Project Management & Project Leadership — University of California, Berkeley (Extension) | PMI Accredited (Completed)

Bachelor of Engineering in Information Technology — Ahmedabad Institute of Technology (2009-2013)

Certified Scrum Master (CSM) — Scrum Alliance (2018, Renewed 2023)

Agent Skills with Anthropic — DeepLearning.AI (May 2026)

Docker Foundations Professional Certificate — Docker, Inc (2026)

Claude Code 4: Agentic Coding for Professional Developers — Anthropic (2025)