

Gaurav Modi

604-716-7917 | g.s.modi@outlook.com | linkedin.com/in/GauravSModi | github.com/GauravSModi

Summary

Software developer specializing in Python backends and full-stack applications, with hands-on production experience in FastAPI, React, cloud deployment, and modern CI/CD workflows.

Technical Skills

Languages: Python, JavaScript/TypeScript, Java, HTML, CSS

Databases: MySQL, PostgreSQL, SQLite, MongoDB, Redis, DynamoDB, Firestore

Frameworks & Libraries: Node.js, Express.js, React, Next.js, FastAPI, Jest, JUnit

Developer Tools: AWS, GCP, Git, GitHub Actions, Docker, NGINX, RabbitMQ, OpenTelemetry, CloudWatch

AI & ML: LangChain, LangGraph, LangSmith, ChromaDB, OpenSearch, AWS Bedrock, Azure AI Foundry

Certifications: AWS Certified Cloud Practitioner (CLF-C02), Microsoft Certified: Azure AI Fundamentals (AI 900)

Experience

Software Developer

May 2025 – Present

Lightning Minds

- Built a LangGraph router that infers clinical role and dispatches queries across FHIR, X12, and CDISC healthcare datasets, achieving 3.8s P50 end-to-end latency
- Deployed a serverless 311 support agent on AWS Bedrock AgentCore with OAuth2-based MCP tool integration
- Built RAG layer on ChromaDB (EC2-hosted) — embedding pipeline + retrieval tuning — reaching 0.92 answer relevance on LangSmith evals
- Developed an 8-stage e-commerce AI agent using LangGraph and FastAPI, achieving 96% task completion rate across 150+ LangSmith eval runs
- Engineered an agent-driven React UI with a dynamic checkout panel, implementing real-time state management driven by LangGraph agent decisions to programmatically control UI visibility, reducing clicks-to-purchase
- Instrumented agent pipeline with OpenTelemetry on CloudWatch for distributed tracing across LLM calls, tool invocations, and retrieval

AI Model Evaluator

Jun. 2024 – May 2025

DataAnnotation

- Evaluated LLM outputs across code generation and reasoning tasks; developed systematic rubrics for identifying failure modes that informed model training improvements across multiple languages and task types

Projects

E-commerce Backend | *Node.js, Express.js, TypeScript, MongoDB, Docker, NGINX, RabbitMQ, Jest*

- Engineered a scalable REST API backend, reducing average response times by 30% during peak traffic via an NGINX load balancer
- Implemented an order processing pipeline, reducing internal processing time by 45% at peak usage
- Covered the REST routes and order-processing pipeline with Jest unit and integration tests at 91% coverage

Clipd — Clipboard history manager for macOS 🍎 | *C++, Swift, CMake*

- Built end-to-end, with a C++ core owning storage, search, and eviction behind a flat extern "C" API and a Swift shell for the menu bar, pasteboard polling, and hotkeys
- Wrote a fuzzy-search scorer that weighs contiguity, word-boundary hits, and recency. Sub-millisecond lookups across 10k+ entries, zero findings under AddressSanitizer and UBSan

Workout-Buddy — Personal AI Trainer 🏋️ | *React.js, TypeScript, FastAPI, Python, SQLite*

- Built an AI personal trainer web app with real-time form tracking and automatic rep counting via a joint-angle state machine on MediaPipe BlazePose keypoints
- Integrated Anthropic Claude API over streaming SSE to generate personalized workout plans and drive hands-free voice coaching via Web Speech API, backed by a FastAPI + SQLAlchemy service for persistent workout history

Education

Simon Fraser University

Bachelor of Science in Computer Science

Burnaby, BC

Graduated Jun. 2024