

Ankit Mukherjee

+1 (716) 709-1597 | ank26.m@gmail.com | LinkedIn | GitHub | Portfolio | San Jose, CA

SUMMARY

Full-stack Software Engineer with 3+ years of experience building high-scale distributed systems and enterprise AI platforms. Proven track record of extreme ownership in fast-paced startup environments, shipping 0-to-1 products and architecting infrastructure that secured a \$12M+ client contract. Expert in Python, TypeScript, AWS, and cloud-native microservices.

EXPERIENCE

Software Engineer – Tech Lead

Jan. 2026 – Present

AskTuring.AI

San Jose, CA

- Partnered with the CEO to lead the Slack integration team, shipping an omni-channel AI chat experience (**12 API endpoints, 5 async workers**) enabling enterprise clients to query RAG-powered knowledge and ingest **100K+ messages** into **Weaviate**.
- Designed **6 PostgreSQL tables** (Alembic, JSONB, composite keys) and engineered Slack event processing with HMAC-SHA256 verification and **Redis deduplication** (exactly-once semantics) to enforce multi-tenant isolation and achieve **99.9% reliability**.
- Enhanced a **44-node LangGraph** RAG pipeline with **hybrid vector+BM25 search, RSE segment reconstruction**, and multi-hop decomposition, optimizing TTFT by **40%** by merging query classification and generation into a single LLM call.
- Built an **LLM-as-judge** evaluation framework with 4 validators and **RAGAS metrics** (faithfulness, relevancy, context precision) using cross-model ensemble judging with bias mitigation, improving RAG answer quality by **30%**.
- Containerized backend services via **Docker** for **Azure Kubernetes Service (AKS)** and deployed an observability stack (**OpenTelemetry, Grafana, Langfuse, PostHog**), reducing mean debug time by **60%** and feature regressions by **40%**.
- Boosted page load speeds by **25%** by eliminating redundant API calls across a **NestJs/TypeScript** architecture, and implemented **SSE streaming** to render AI responses in real-time.
- Served as primary **code reviewer** for full-stack PRs (FastAPI/Next.js), catching **15%** of bugs pre-merge, and accelerated team onboarding by **50%** by building custom **Claude Code** developer workflows.

Software Engineer – High-Scale Systems

July 2021 – July 2024

PwC

Kolkata, India

- Spearheaded a GenAI initiative that secured a **\$12M+ client contract**, architecting a **RAG** pipeline using **AWS Bedrock Knowledge Base** and **AWS Lambda** functions to intelligently resolve complex fallback queries.
- Partnered directly with a Director to serve as the **Product Manager** and Tech Lead for an internal enterprise platform, translating complex business requirements into scalable technical architectures adopted by **1,000+ users** across **5+ major engagements**.
- Built and scaled containerized **REST API** microservices (**Node.js, Docker, AWS ECS**) with **Kafka** messaging and **DynamoDB** GSIs, reducing latency by **40%** and seamlessly sustaining **300% traffic spikes**.
- Modernized deployments by provisioning Infrastructure as Code (IaC) via **AWS CloudFormation**, building automated **CI/CD** pipelines in **GitHub Actions**, and implementing **AWS CloudWatch** observability to maintain **99.99% availability**.
- Engineered a secure revenue engine using **Stripe APIs** processing **\$1M+ annually**, integrated into a highly performant **React/TypeScript** frontend with **85% test coverage** via Jest and Cypress.

PROJECTS

Founding Engineer – Filmic Technologies | *Nest.js, React, AWS, PostgreSQL*

Seed-Funded Startup

- Architected the **0-to-1 digital film archive marketplace** using **Nest.js**, enabling successful **seed funding**.
- Engineered a real-time messaging system with **WebSockets** and designed a normalized **PostgreSQL** schema for complex transactions.

Medical Imaging Classification | *GitHub | PyTorch, CNNs, Deep Learning*

Computer Vision

- Developed and trained **CNNs** on the **BreastMNIST** dataset with **PyTorch**, achieving strong clinical classification accuracy.
- Evaluated model robustness using **confusion matrices, ROC curves, F1 scores**, and systematic **dataset analysis** to improve diagnostic performance.

TECHNICAL SKILLS

Languages: Python (FastAPI), Node.js (NestJS), TypeScript, Swift, C++, SQL, Java, Go (Basic)

GenAI & ML: LLMs (OpenAI, Gemini, Anthropic), LangGraph, Agentic RAG, PyTorch, CNNs, NLP, Computer Vision

Data & Infrastructure: Weaviate, pgvector, Redis, Kafka, Dramatiq, DynamoDB, PostgreSQL

Cloud & DevOps: AWS (EKS, Lambda, Bedrock, CloudFormation), Docker, Kubernetes, GitHub Actions, CI/CD

Frontend & Mobile: React, Redux, Next.js, iOS Development, JavaScript, HTML5/CSS3

Observability & Evaluation: Prometheus, Grafana, Langfuse, PostHog, RAGAS

Concepts: Distributed Systems, Microservices, Event-Driven Architecture, Data Structures & Algorithms, REST APIs

EDUCATION

University at Buffalo, SUNY

Buffalo, NY

Master of Science in Computer Science; GPA: 3.73

Dec. 2025

Institute of Engineering and Management

Kolkata, India

Bachelor of Science in Information Science

July 2021