

Abhay Shah

shahabhay941@gmail.com | [linkedin.com/abhay](https://www.linkedin.com/in/abhay) | github.com/abhay | Mumbai, India

EDUCATION

Columbia University in the City of New York

Master of Science, Applied Analytics | GPA: 3.7 / 4.0

New York, NY

Dec 2024

University of Mumbai

Bachelor of Science, Computer Science | GPA: 9.7 / 10.0

Mumbai, India

May 2021

LANGUAGES AND SKILLS

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

Tools/Frameworks: React.js, Node.js, MongoDB, MySQL, Express.js, Next.js, PostgreSQL, Supabase, Spring Boot, REST API, Jenkins, Git, LangChain, Large Language Models

Cloud/Devops: AWS Lambda, S3, EC2, RDS; GCP, Docker, Kubernetes, CI/CD, Github Actions, Terraform

EXPERIENCE

Integrow Asset Management

March 2025 – Present

FullStack AI Engineer & Strategy (Founder's Office)

Mumbai, Maharashtra

- Evaluated Annual Operating Plan for a \$50M real estate credit fund raise; stress-tested quarterly drawdown targets and flagged critical errors ahead of board presentation.
- Developed Intella, a **RAG-based AI assistant** powered by AWS Bedrock, LangChain, FastAPI, and PostgreSQL that integrated multiple internal data sources, improving research and **knowledge access**.
- Architected XEPRA, a digital real-estate wallet platform enabling investors to transact and manage alternative investments with **\$54M assets** mapped in the pilot phase.
- Built and deployed RIS (Research Intelligence Summary) automation using HubSpot API with a FastAPI backend on Render, handling **more than 1K internal requests** daily for the C-Suite.
- Built and deployed MaaS, an asset monitoring platform on AWS RDS, Docker, Kubernetes, providing real-time oversight of AIF assets; **reduced manual tracking by 60%** and **cut reporting turnaround time by 40%**.

Turmerik

Sept 2024 – Feb 2025

Software Engineer - Machine Learning

New York, NY

- Built Turmerik Terminal, a healthcare SaaS that analyzes NIH clinical studies and scores their investment potential, like a Bloomberg Terminal for clinical trials; enabled faster due diligence, **reducing research time by 40%**.
- Developed an **XGBoost ML model** that generates a **Turmerik Score**, improving investment signal accuracy compared to heuristic methods.
- Integrated NIH API and ClinicalTrials.gov API to ingest and process over **25,000+ trials**, ensuring real-time, scalable insights for biotech investors.
- Contributed to securing a **\$4M Grant** from the National Institute of Health.

Tata Consultancy Services - Digital

July 2021 – Sept 2022

Software Engineer - Backend

Mumbai, India

- Reduced **processing times for workflows by 40%** by developing REST APIs using **Spring Boot, Kafka, and AWS**, ensuring seamless data exchange through test-driven development.
- Improved DB query performance by 10% via Hibernate-based backend and Docker/Kubernetes scaling.
- Enhanced API throughput by **25% via payload compression**, JUnit regression tests, and Mockito-based automated test coverage, improving system reliability during peak loads
- Implemented CI/CD pipelines with Jenkins and Docker, **cutting release cycles by 15%** and reducing deployment errors, enabling faster delivery for multi-country digital transformation projects.

PROJECTS

HealthLlama - Fine-tuning LLMs with Kubernetes | Python, GCP, Docker, Kubernetes, Llama3, LLM

- Fine-tuned LLaMA-3 (8B) on the HealthCareMagic 100K dataset using QLoRA and Hugging Face TRL, **achieving a 15% BLEU score** and **+0.01 BERTScore** improvement on healthcare-specific NLP tasks.
- Containerized and deployed the model on **GCP with Docker and Kubernetes**, serving an interactive Streamlit chatbot that handled concurrent inference requests with 3× replication scaling.
- Implemented experiment tracking with Weights Biases and optimized inference with quantization (4-bit), reducing memory usage by **40% while maintaining accuracy**.