

Affan Arif Khamse

Backend Software Engineer | LLM Systems | Scalable Architectures | NYU CS Grad 2025

New York, NY · (516) 853-4972 · khamseaffan@gmail.com · github.com/khamseaffan · linkedin.com/in/affan-khamse

Education

Master of Science in Computer Science (GPA: 3.78/4)

New York University

Courses: Software Engineering, Data Structures & Algorithms, Machine Learning, Databases

New York, USA

Expected May 2025

Bachelor of Engineering in Computer Engineering (GPA: 9.00/10)

University of Mumbai

Mumbai, India

June 2023

Technical Skills

Programming and Scripting Languages: C++, Java, Python, Node.js, JavaScript

Frameworks and Libraries: Flask, Spring Boot, FastAPI (RESTful), Django, ReactJS

Cloud and DevOps: Amazon Web Services / AWS, Azure, Docker, Kubernetes, GitHub Actions, Travis CI

Databases & Query Languages: SQL(MySQL, PostgreSQL), NoSQL(MongoDB, Firebase)

Testing & Debugging: JUnit, PyTest, Django Test, Chrome DevTools

Tools and Methodologies: Git, GitHub, Figma, UML, Agile (Scrum), Object-Oriented Design

Specialized Expertise: API Design, Microservices Architecture, LLM Applications, Scalable Systems

Work Experience

Software Engineer

InquisAI (NYU R&D Project) | inquis-ai.com

New York, USA

Jun 2024 – Present

- **Redesigned backend architecture**, migrating from Flask to **FastAPI** with async I/O, reducing latency by 30% and boosting throughput for high-concurrency production queries
- Developed **scalable RESTful APIs** for vector-based retrieval and deployed across Linux-based infra to support 1K+ concurrent users
- Led infra decisions and deployment strategy during Agile sprints, leveraging Azure DevOps for CI/CD and monitoring
- Integrated **OpenAI embeddings** and LangChain with vector stores for performant document Q&A across distributed sources

Software Teaching Assistant – Object-Oriented Programming

Courant Institute of Mathematical Sciences @ New York University

New York, USA

Jan 2024 – Present

- Mentored 50+ students in **C++ and Java**, emphasizing object-oriented principles, algorithmic efficiency, and memory management
- Conducted detailed **code reviews** and debugging walkthroughs to reinforce production-level software practices

Projects

Home Store – E-Commerce Platform (Backend)

Jan 2025 - Present

Spring Boot, Spring Cloud, Docker, PostgreSQL, React Router, Firebase, Microservices, SwaggerUI

[GitHub](#)

- Built containerized microservices on **Spring Boot**, integrated via **Eureka service discovery** for modular scalability and fault isolation
- Designed and documented **RESTful APIs** for order/user/cart workflows, integrating seamlessly with frontend and backend services
- Emulated production environments using Docker Compose, enabling robust multi-service testing with **PostgreSQL backend**

FlashBids - Full Stack

Sep 2024 - Dec 2024

AWS EC2, DynamoDB, S3, Redis, WebSockets, CloudWatch, Flask, Python, REST APIs, DevOps

[Demo](#) | [GitHub](#)

- Engineered a real-time auctioning system handling **20K+ users** using EC2 Auto Scaling, Redis caching, and DynamoDB for resilient writes
- Implemented a **low-latency WebSocket layer** for live bid updates, optimizing client-server communication and reducing lag by 40%
- Built non-blocking bidding APIs with CloudWatch instrumentation to monitor throughput and response spikes under load
- Automated an event-driven bid pipeline using **Amazon SQS** and SES, ensuring 99% delivery across distributed event chains

VibeCheck (Full-Stack Project)

Sep 2023 - Dec 2023

Django, Python, Bootstrap, AWS Elastic Beanstalk, PostgreSQL, Redis, TravisCI

[Demo](#) | [GitHub](#)

- Developed a full-stack social matching platform leveraging real-time Spotify data and **Redis Pub/Sub** for event-driven messaging
- Deployed on **Linux-based Elastic Beanstalk** with CI/CD via TravisCI and integrated Postgres for structured data storage
- Achieved 87% code coverage through unit + integration tests, ensuring stability across rapid release cycles