

Shikhar Malik

shikhar.malik@stern.nyu.edu

My Website

Linkedin

Github

Brooklyn, NY

EDUCATION

New York University (NYU)

Master of Science - Computer Science, GPA: 3.86

Brooklyn, NY

Sept 2024 - May 2026

Core Coursework: Deep Learning, Machine Learning, Big Data, Data Science, Information Visualization

Management Coursework: Digital Strategy, Healthcare Transformation & Emerging Technologies

Delhi Technological University

Bachelor of Technology - Computer Science

New Delhi, India

Aug 2017 - May 2021

Coursework: Software Engineering, Object Oriented Programming, Database Management, Data Structures & Algorithms

WORK EXPERIENCE

Root Access

Software Engineer Intern

New York City, US

June 2025 - August 2025

- Worked in a fast-paced, pre-seed startup environment, contributing across the full stack (TypeScript, React, FastAPI, AI).
- Implemented secure authentication (OAuth, Auth0) and optimized LLM prompt engineering enhancing the model's code, hardware-specific code generation accuracy and contextual understanding.
- Led development of the company's flagship VS Code-based AI editor, core product contributions directly supported a successful \$2.1M seed funding round.

Optum (United Health Group)

Associate Software Engineer

Gurugram, India

June 2021 - August 2024

- Owned high-volume applications processing over 27 million records each quarter, contributing to a \$7M cost savings in the company's provider paperless initiative in healthcare infrastructure domain.
- Automated critical reports and ETL workflows by developing Python scripts & Java schedulers with MongoDB integration, reducing manual effort, improving processing efficiency and saving the team 15+ hours per week while enhancing operational visibility.
- Resolved a critical memory leak, reducing system restarts from 3+ per day to 0, enhancing processing stability for high-load operations, improving API performance by 100%.
- Migrated critical applications from OpenShift to on-prem Kubernetes, improving system stability from 96% to 99% uptime and implemented CI/CD pipelines using Jenkins, streamlining deployment workflows and automating release processes.
- Served as an SRE and on-call engineer, took full ownership of 50+ end to end production releases ensuring stable deployment for millions of users and handling high priority incidents and war rooms under strict SLAs.
- Received 14+ Bravo Awards and High Flyer Award in 2022 for outstanding contributions to automation, modernization, optimization and leadership.

PROJECTS

Collabdesk - Cloud-Native Collaborative Workspace Platform

[[Link](#)]

- Developed a full-stack collaborative workspace platform using React 19 and Django REST Framework, implementing Auth0 for multi-tenant authentication and custom middleware to enforce strict workspace-isolated user access control.
- Integrated Google Generative AI API to enable intelligent document parsing (PDF/DOCX) and multi-turn chat persistence, leveraging AWS S3 and Boto3 to deliver secure file management and real-time editable system notes.
- Architected a multi-environment AWS cloud infrastructure (Elastic Beanstalk, S3, CloudFront) secured by end-to-end HTTPS encryption via ACM, automated through a Travis CI pipeline to streamline branch-based production deployments.

StyleSnap - NPU-Accelerated Edge AI Virtual Fashion Assistant

[[Link](#)]

- Developed a local-first virtual fashion desktop application using React, Tauri, and FastAPI, implementing client-side image processing and background removal via Hugging Face Transformers to ensure complete user data privacy.
- Integrated OpenAI CLIP (ONNX) and Llama 3.1 (8B) models to automatically analyze, tag, and parse natural language queries, generating context-aware outfit recommendations stored in a local SQLite database.
- Optimized a Stable Diffusion try-on pipeline leveraging the Qualcomm QNN/QAIRT SDK on a Snapdragon X Elite NPU, accelerating inference runtime to 5-7 seconds per image to achieve a 5x speedup over CPU fallback.

DermaVision - Skin Cancer Detection Toolkit

[[Link](#)]

- Won 2nd place in the Optum Global Hackathon out of 300+ global teams, developing a full-stack web application for skin cancer detection with real-time image-based predictions using deep learning across 7 diagnostic categories.
- Built a Flask-based computer vision microservice with a DenseNet169 classifier trained on the HAM10000 dataset, achieving 91.2% accuracy. Implemented user authentication system with SpringBoot microservices and a React frontend.

SKILLS

- Languages:** TypeScript, JavaScript, Python, Java, C++, HTML, CSS, SQL
- Frontend Development:** React, Vite, Tailwind CSS, Tauri, Electron
- Backend Frameworks:** FastAPI, Flask, Django, Spring Boot, Node.js, REST
- Databases & ORM:** MongoDB, MySQL, Supabase, SQLite, Spring Data JPA
- DevOps & Cloud:** AWS (Elastic Beanstalk, S3, CloudFront), Docker, Kubernetes, Jenkins, Git, Linux, Github Actions
- Testing & QA:** JUnit, JMeter, SonarQube, CodeQL, Postman, Agile Methodologies, Jira, ServiceNow, Grafana
- ML and AI:** Tensorflow, Pytorch, Hugging Face, LLMs, RAG, Prompt Engineering, Agentic AI