

SUBHAN QURESHI

AI/ML Engineer | Full-Stack Developer

+92-314-0230083 | subhan.queshi1915@gmail.com | Karachi, Pakistan | <https://github.com/SUBHANQURESHIO>

PROFESSIONAL SUMMARY

Final-year Computer Engineering student and group leader of an AI SaaS final-year project: a multi-agent LLM platform built on LangGraph, RAG and self-hosted vLLM inference. Prior intern at Payfast building authentication and role-based access control on a live backoffice portal. Seeking an AI or software engineering internship in Karachi or remote.

EDUCATION

B.E. Computer Engineering — SSUET, Karachi

2022 – 2026 (expected)

Sir Syed University of Engineering & Technology | Batch 2022F

FINAL YEAR PROJECT

The AI Dietician — Group Leader

2025 – 2026

SSUET | Supervised by Miss Maheen Danish

Cloud-native, multi-agent SaaS delivering mathematically audited, clinically validated meal plans across 9 conditions and 4 athlete profiles. Led architecture and full-stack development.

- Designed a 4-agent LangGraph pipeline (Intake → RAG Researcher → Diet Planner → Clinical Critic) where the critic audits every plan against hard calorie/protein tolerances and rejects failures for regeneration.
- Built the FastAPI backend and orchestrator; served self-hosted Llama 3.1 8B via vLLM on cloud GPU, with async 202 + WebSocket streaming to avoid proxy timeouts.
- Implemented two-stage Pinecone retrieval (bi-encoder recall + cross-encoder rerank) over USDA, WHO/ADA, ISSN and Pakistani datasets: 120+ local foods, 200+ clinical rules.
- Shipped a Next.js 15 web client and React Native (Expo) app with offline caching and Twilio WhatsApp check-ins; authored the IEEE paper and SRS.

Stack: Python, FastAPI, LangGraph, Llama 3.1, vLLM, Pinecone, PostgreSQL, Next.js 15, React Native/Expo, Twilio, Docker, Render, Vercel

WORK EXPERIENCE

Software Development Intern — Payfast

Aug 2024 – Oct 2024

- Built frontend and backend for the internal backoffice portal (Next.js, MongoDB) inside a live production codebase.
- Implemented secure auth, email flows and role-based access control using NextAuth, JWT, Bcrypt.js and Nodemailer.

Freelance Full-Stack Developer — E-Commerce Site

- Built a Next.js storefront with Sanity.io CMS, dynamic product pages, cart and checkout.

Private Tutor — Self-employed

2024 – Present

- Teach mathematics and core subjects to junior secondary students with structured lesson plans.

PERSONAL PROJECTS

- News platform (Next.js + Sanity.io) with content management and dynamic routing.
- Full-stack Todo REST API: FastAPI, PostgreSQL and a Next.js frontend with CRUD.

TECHNICAL SKILLS

Frontend: Next.js, React, TypeScript, JavaScript, Tailwind CSS, Shadcn/ui

Backend: Node.js, FastAPI, REST APIs, JWT Auth, WebSockets

AI & ML: LangGraph, LangChain, Pinecone (RAG), vLLM, Llama 3.1, Groq API, Prompt Engineering

Data & Cloud: PostgreSQL, MongoDB, Prisma, Sanity.io, Docker, Render, Vercel, RunPod/Vast.ai, Git

ADDITIONAL

- PIAIC — Generative AI, Web3 & Metaverse: completed 3 of 5 quarters.