

Skills

- Experienced in developing and deploying **AI-powered applications** using **FastAPI, LangChain, and React**, integrating **LLMs** for conversational intelligence, content generation, and workflow automation.
- Strong command of **Python, Node.js (Express 5), and React (Vite)**, applying **clean architecture** and modular 3-layer design principles for scalable backend and frontend systems.
- Built production-grade **multi-tenant AI platforms** with **Prisma ORM, PostgreSQL/MySQL, Redis/Bull queues**, and **Cloudinary**, supporting real-time media pipelines and AI-driven property listing generation.
- Skilled in **MLOps and deployment workflows**, including **Docker-based environments**, CI/CD, secure API management, and deployment on **Railway, Heroku, and AWS**.
- Experienced in **vector search integration** and **LLM tool orchestration**, enhancing model performance through memory handling, fallback models, and caching for latency reduction.
- Focused on **secure AI system design**, implementing JWT authentication, rate limiting, XSS prevention, signed URLs, and session-bound validation to ensure robust application integrity.
- Collaborative communicator and technical lead, mentoring peers in **prompt engineering, backend architecture**, and **AI-enhanced application development** aligned with production standards.

Work Experience

Ai Developer – Forrentech

Lahore, Pakistan | Apr 2025 – Present

- Served as Lead Developer in a 3-member team for an AI-powered property listing application that auto-generates listings from uploaded property images.
- Engineered a robust backend using Node.js (Express 5) with a 3-layer architecture, leveraging Prisma ORM and PostgreSQL for clean, scalable data flow and modular structure.
- Built a performant Vite + React frontend with seamless social media integration and export-to-portal features, enhancing listing visibility and marketing automation.
- Deployed the system via Docker, using Heroku for staging and AWS for production, ensuring a smooth CI/CD pipeline.
- Optimized AI image processing and API workflows, improving generation speed by 35% and reducing API costs by over 60% through streamlined requests and efficient caching.

Chatbot developer – Optivex solutions

Lahore, Pakistan | Jan 2023 - Mar 2025

- Served as Associate Developer for a medical AI chatbot that provides real-time assistance and emergency response guidance through intelligent conversational flows.

- Engineered a modular FastAPI backend integrated with a Vite + React frontend, ensuring seamless communication via robust API integration.
 - Utilized MySQL for structured data management and user interaction logging, supporting reliable and scalable chatbot analytics.
 - Deployed the system on Railway, achieving high availability and smooth CI/CD pipelines for continuous delivery.
 - Optimized the response pipeline to deliver 100–500 ms response times, ensuring rapid guidance during medical emergencies.
-

Projects

Realestic-Ai (End to End solution for Real Estate Listings)

- Built a multi-tenant real-estate platform (Portal + Agencies) with Node.js/Express 5, Prisma/Postgres, React 19/Router 7, Tailwind, Radix UI, Recharts, Redis/Bull, Cloudinary, and OpenRouter.
- Delivered AI-powered listing generation with a resilient 3-step flow: signed client-side Cloudinary uploads → AI processing (GPT-4o with fallback to 4o-mini) → finalize + atomic image moves; eliminated 30s timeouts and achieved sub-30s end-to-end for 15 images.
- Secured the media pipeline with signed public IDs, folder scoping, session-bound URL validation, extension whitelisting, and 24h temp cleanup, preventing cross-session tampering.
- Built CSV import + AI enrichment queues with Redis/Bull (batch size 7, 3 retries, backoff, per-batch progress, error CSV output) and persisted progress across navigation; upserted ImportedListings and promoted to Listing on enrichment.
- Shipped portal/agency dashboards with KPI cards, Implemented quota tracking (per-portal and per-agency) with accurate “Listings This Month” metrics counting all created + enriched-this-month entries and clear progress indicators.
- Unified export UX (CSV + CMS) for portals and agencies with consistent response shapes, dropdown actions, and toasts (loading/success/failure) via react-hot-toast + Toaster at root.
- Hardened auth and scoping with JWT, portal/agency middleware, rate-limit and axios token interceptors with auto-logout on 401.
- Improved reliability of AI services with model fallback, structured logging, and graceful degradation when providers return 4xx/5xx; added input validation
- Optimized DB access with parallel queries (Promise.all), aggregation via Prisma groupBy for rankings/stats, and atomic transactions for finalize/enrichment flows.

Autonomous Healthcare Risk Assistant (GenAI)

- Engineered a predictive XGBoost model on a 1M-row dataset with 5-fold cross-validation, automating risk scoring for 200K test patients, reducing readmissions by 15%, saving \$3.5M annually.
- Developed a real-time Dash dashboard with a Flask API, enabling 30-second updates for 6 visualizations across 1K daily patient queries, cutting intervention delays by 30%, saving \$1.2M in operational costs.
- Analyzed temporal trends with Plotly, identifying a 15% winter readmission spike in 61-80 age group across 600K records, optimizing resource allocation to lower peak-season readmissions by 10%, saving \$2M.

Real-Time Omani Therapist Voice (GenAI Voice Chatbot)

- Built a culturally-adapted mental health assistant for Arabic-speaking users using GPT-4, Azure STT/TTS, and emotion-aware response logic.
 - Implemented a modular voice-only pipeline using Flask, real-time transcription, emotion detection (BERT-based), GPT-4o therapy agent, and WAV-based TTS synthesis.
 - Achieved <20s latency and 96% satisfaction in demo testing, with embedded safety filters for suicide/harm detection and dialect normalization.
 - Packaged project with local fallbacks (offline model storage, local logs, cultural safety gates), designed for zero-cost deployment and long-term reliability.
-

Education

Bachelors in Artificial Intelligence

Superior University, Lahore | 2022 - 2026 (completing while working in parallel)

- Relevant Courses: Data Science, Database Systems, Python Programming, Statistics.
 - Certifications: Ai Engineer for DS Associate by DataCamp , Supervised learning , Machine learning , Cloud Computing.
-
-