

Saqib Ali Shah

SOFTWARE ENGINEER

+92 347-5287822 | saqibshah2882@gmail.com | saqib-002.github.io | linkedin.com/in/saqibshah172 | github.com/Saqib-002

SUMMARY

Software Engineer with a proven track record of architecting scalable full-stack systems and high-performance Agentic RAG workflows. Specialized in engineering intelligent infrastructure, such as B2B lead platforms and secure natural language database interfaces, using FastAPI, React, and Gemini Vertex AI. Focused on transforming complex data bottlenecks into production-ready, AI-driven products that deliver measurable cost savings and sub-second performance.

EDUCATION

Institute of Space & Technology

Bachelor of Science in Computer Science

Islamabad, Pakistan

Sep 2020 – July 2024

- **Gold Medalist:** Awarded for securing 1st position in the graduating batch for academic excellence.

KRL Model College For Boys

HSSC (Pre-engineering)

Kahuta, Pakistan

2018 – 2020

EXPERIENCE

Software Engineer

Modern Intelligent Solutions

Jul 2024 – Dec 2025

Riyadh, Saudi Arabia

- Architected "Imagini," a high-performance B2B leads platform housing millions of records; engineered a credit-based access system and streaming upload pipelines using FastAPI and React to ensure sub-second search responses.
- Engineered a multi-agent RAG framework to navigate 1.5M Arabic documents, implementing hierarchical agent workflows and graph-based traversal to handle sophisticated linguistic queries with high precision
- Engineered a secure natural language database interface for the Saline Water Conversion Corporation (SWCC) using LangGraph and Gemini, implementing strict intent classification to prevent unauthorized data modification.
- Designed an RFM-based customer segmentation engine for premium airport services, deploying predictive models to automate dynamic subscription plans and reduce operational costs.

Software Engineer

Voilabs

Nov 2023 – Feb 2024

France (Freelance)

- Fine-tuned DONUT transformer models for OCR-free extraction of French bill data, optimizing the architecture for mobile deployment via ONNX conversion.
- Integrated transformer-based machine learning models into Next.js applications, creating reusable components that improved AI feature integration speed.
- Built robust data processing pipelines using Python, Pandas, and NumPy, and fine-tuned neural networks to improve specific business use-case accuracy by 25%.

ML Engineer Intern

Ksoft

Aug 2023 – Sep 2023

Islamabad, Pakistan

- Implemented computer vision solutions using YOLO for real-time object detection tasks.
- Designed intuitive user interfaces using Figma and contributed to Mobile App development with Kotlin.

TECHNICAL PROJECTS

ThunderJaw Restaurant POS | Electron, React, TypeScript, PostgreSQL, Knex

- Developed a cross-platform desktop application for restaurant management featuring real-time order tracking, kitchen management, and automated PostgreSQL database migrations.
- Implemented secure role-based access control (RBAC) and modular state management to handle complex inventory, worker salaries, and delivery logistics in high-traffic environments.

High-Scale Email Verification System | Python, GCP, Docker, Celery, FastAPI

- Engineered a multi-layered validation pipeline (DNS resolution, SMTP probing, and heuristics) that reduced verification costs by 90% compared to commercial providers.
- Implemented asynchronous workers and distributed task queues to process hundreds of thousands of emails concurrently while avoiding IP blacklisting.

Mandumah Multi-Doc Agent | LlamaIndex, Gemini Vertex AI, FastAPI, Python

- Developed an agentic RAG system to navigate 1.5M Arabic academic documents using a hierarchical meta-agent architecture and graph-based traversal.
- Optimized retrieval for complex Arabic linguistic nuances, leveraging tool-calling for automated document summarization and deep-dive query answering.

B2B Leads Platform (Imagini) | React, FastAPI, PostgreSQL, Redis, Celery

- Architected a scalable platform managing millions of leads, featuring streaming data ingestion and a credit-based access system for B2B users.
- Optimized database indexing and caching to maintain sub-second search and filtering performance across massive datasets.

CERTIFICATIONS

Generative AI Fundamentals | *Google*

AI For Everyone | *Coursera*

Pandas | *Kaggle*

Applied Software Engineering Fundamentals | *Coursera*

Linear Algebra for Machine Learning and Data Science | *Coursera*

Calculus for Machine Learning and Data Science | *Coursera*

Backend Development & DevOps | *Boot.dev*

TECHNICAL SKILLS

Languages: JavaScript (ES6+), TypeScript, Python, SQL, HTML5, CSS3, Go

Frontend & Desktop: React.js, Next.js, Redux, TailwindCSS, Material UI, Electron

Backend & Cloud: Node.js, FastAPI, Express.js, Docker, Flask, AWS, Google Cloud Platform (GCP)

AI & Machine Learning: PyTorch, LangChain, LangGraph, Gemini Vertex AI, RAG Pipelines

Databases & Tools: MongoDB, PostgreSQL, Redis, Supabase, Pinecone, Weaviate, Knex.js (ORM)