

TALHA JALEEL

+92-317-7201520 | talhajaleel2@gmail.com | linkedin.com/in/talhajaleel

Pakistan (Open to Remote Roles Globally)

PROFESSIONAL SUMMARY

Senior Software Engineer and AI/ML Engineer with 6+ years of experience building scalable, cloud-native full stack applications. Expert in Python (Django, FastAPI), JavaScript (React, Node.js, TypeScript), and AWS cloud infrastructure. Specialized in MLOps, LLM integration, RAG pipelines, and production AI systems. Proven track record in backend development, microservices architecture, and responsive frontend interfaces. Strong experience with CI/CD pipelines, containerization (Docker, Kubernetes), and agile development methodologies.

TECHNICAL SKILLS

Languages: Python, JavaScript, TypeScript, SQL, HTML, CSS

Frontend: React, Next.js, Redux, Tailwind CSS, Responsive Design, REST APIs, GraphQL

Backend: Django, Django REST Framework, FastAPI, Flask, Node.js, Express.js, Microservices, API Development

AI/ML & MLOps: LLMs, RAG, LangChain, GPT-4, Azure OpenAI, Prompt Engineering, Neural Networks, Computer Vision, NLP, TensorFlow, PyTorch, Model Training, Model Deployment, Embeddings

Databases: PostgreSQL, MongoDB, Redis, Pinecone (Vector DB), SQL, NoSQL, Database Design

Cloud & DevOps: AWS (EC2, S3, Lambda), Azure, Docker, Kubernetes, CI/CD, GitHub Actions, Linux/Unix, Git

Testing & Tools: Unit Testing, Jest, Pytest, Selenium, Playwright, Agile, Scrum, System Architecture

PROFESSIONAL EXPERIENCE

Software Engineer | Nuclieos | Apr 2024 – Present

- Developed production LLM-based chatbots using RAG pipelines, achieving 35% improvement in response accuracy through model training and fine-tuning
- Designed and deployed scalable RESTful APIs using Django, FastAPI, and Node.js/Express.js for real-time AI application integration
- Built MLOps pipelines for model deployment and monitoring, reducing deployment time by 40% using Docker and Kubernetes
- Implemented CI/CD pipelines with GitHub Actions, achieving 99.5% deployment success rate with comprehensive unit testing
- Leveraged AWS services (EC2, S3, Lambda) and Linux/Unix environments to scale backend services handling 50K+ daily requests

Machine Learning Engineer | ESS | Apr 2023 – Mar 2024

- Built Sales AI Voice Bot MVP using LLaMA, neural networks, and NLP, integrating Twilio for telephony and Pinecone VectorDB for semantic search
- Developed real-time computer vision system using Python, integrating LIDAR and camera data for object detection at 30fps
- Implemented backend services in Node.js and Django with MongoDB and Redis caching, reducing API latency by 45%
- Deployed ML models on AWS with Docker and Kubernetes, establishing MLOps best practices for reproducible model training and deployment
- Wrote comprehensive unit tests using Pytest and Jest, achieving 85% code coverage across ML and backend services

Full Stack Developer | Ilsa Interactive | Nov 2020 – Apr 2023

- Led full stack development of Fayvo social app (100K+ users, 1K+ DAU) using React, Redux, TypeScript frontend and Django REST Framework backend
- Designed ML-powered Recommendation System using neural networks and PostgreSQL, increasing user engagement by 28%
- Built responsive frontend interfaces with HTML, CSS, and React, ensuring cross-browser compatibility and mobile-first design
- Developed Adult Content Detection System using computer vision and deep learning, achieving 96% accuracy in content moderation
- Managed database design with PostgreSQL and MongoDB, optimizing complex queries and reducing response times by 50%
- Deployed scalable microservices on AWS (EC2, S3) with Docker containers, maintaining 99.9% uptime

PROJECT HIGHLIGHTS

Open Panacea – Full stack healthcare platform using AWS, FastAPI, GraphQL, Next.js with TypeScript. Implemented responsive design and RESTful APIs serving 10K+ patients.

KIRU – AI-powered HR platform for bulk resume analysis using LLMs and neural networks. Built React/Redux frontend and Django backend with MLOps pipeline for model deployment.

Dari – Mobile-first service marketplace with Node.js/Express.js backend, MongoDB database, and Stripe/Tamara payment integration. Deployed on AWS with Kubernetes.

EDUCATION

Bachelor of Science in Computer Science | FAST NUCES, Pakistan

CERTIFICATIONS

- Data Structures in Python – Coursera
- Facial Recognition using Python – RealPython
- JavaScript Algorithms and Data Structures – freeCodeCamp