

Afzal Khan

✉ afzaljawadkhan@gmail.com

☎ +92 337 9755627

in [Afzal Khan](#)

EXPERIENCE

ClaimbAI (Remote)

California, USA

AI Engineer

July 2025 – Present

- Architected & shipped an **agent-first ecosystem** using Google ADK, n8n, Firebase Functions, and Cloud Run powering web + iOS/Android frontends. Built secure connectors for Gmail, Calendar, Drive and task systems enabling agent access to user resources.
- Designed RAG and alternative SLM workflows: evaluated small local models vs. RAG setups and integrated retrieval layers using PostgreSQL vector stores and LlamaIndex-like patterns for knowledge access.
- Implemented production REST/FastAPI backends, orchestrated CI/CD pipelines, and set up monitoring for latency/errors across agent services to ensure reliable deployment and rollback strategies.
- Built automated pipelines (questionnaire → Drive → JSON → recommendations) that feed into UX components; improved end-to-end throughput and observability.
- Collaborated with product, mobile, and security teams to design privacy-preserving connectors and role-based access for agent operations.

COMCEPT

Islamabad, Pakistan

Software Engineer

Jan 2025 – October 2025

- Developed real-time computer vision and tracking stacks (YOLOv9–v11, GOTURN, CSRT, SAMURAI) and deployed on NVIDIA Jetson Orin, achieving
- Optimized models and inference via TensorRT, pruning, and quantization — improved inference speed by ~30% and reduced false positives by 25%.
- Built C++/Qt visualization and monitoring dashboards (40+ screens) and embedded control systems with velocity-aware PID tuning for high-stability tracking.
- Led system integration work across firmware, compute nodes (Radxa/Armbian), and networked devices—reducing boot time and improving system uptime.

AGENIUS

Islamabad, Pakistan

Chief Operating Officer (COO)

July 2024 –December 2024

- Led operations, recruitment, and growth initiatives, balancing executive duties with hands-on technical contributions.
- Directed a major data transformation project: converted a 200M-row dataset across ~200 tables into a normalized 3NF data warehouse and built star-schema semantic layers for analytics and forecasting.
- Developed innovative tech solutions, including a Node.js/React chatbot, a custom web scraping tool, and a proof-of-concept player tracking system for football games using OpenCV and computer vision.
- Built an agent-enabled Node.js/React chatbot and automated ETL using AWS Glue and SQL; simplified analytics workflows and improved dashboard refresh times and UX.
- Managed client relationships and technical deliverables, aligning cross-functional teams (sales, engineering, analytics) to deliver measurable outcomes.

Convergent Business Technologies

Islamabad, Pakistan

Consultant Analytics and Insights

August 2023–July 2024

- Implemented market-basket analysis (Apriori), dimensional modeling, and dashboarding (Tableau, DBeaver). Automated analytics refresh pipelines and reduced operational overhead for multiple clients.
- Conducted ETL to transform raw data into a 3NF data warehouse and developed a dimensional model semantic layer for it.
- Automated the Pepsi KSA & UAE project, reducing the refresh cycle time from 1 week to 3 days (a 57% reduction) and cutting team size from 6 to 2 through custom scripting solutions and query tuning.
- Built ML pipelines for transcription and audio analytics (OCR + timestamped ASR + post-processing) used in aviation pilot-tower prototyping

Upwork

Islamabad, Pakistan

Freelance Data Engineer and Developer

August 2022–August 2023

- Delivered end-to-end analytics and predictive models (churn, forecasting, inventory optimization) for telco, healthcare, and finance clients. Built ETL, dashboards (Power BI / Tableau), and automated reporting systems.
- Built automated ETL workflows, leveraging SQL for database queries and AWS Glue for efficient data pipeline management.
- Conducted statistical analyses to uncover actionable insights, using tools such as pandas, NumPy, and scikit-learn.
- Provided consulting services to optimize data-driven strategies for SMEs, including training teams on analytics tools and dashboard usage

Core Competencies

- **AI / LLMs & Agents:** LangChain, LlamaIndex, Retrieval-Augmented Generation (RAG), local SLMs (Ollama/Self-hosted), Google Agent Development Kit, n8n
- **APIs & Integrations:** OpenAI, Anthropic, Gemini, REST/FastAPI, Zapier/Make, webhook-based integrations, OAuth/connectors for email/calendar/tasks
- **ML / DL:** PyTorch, TensorFlow, Hugging Face, Transformer models, Whisper, Librosa, YOLO (v9–v11)
- **Model Optimization & Edge:** TensorRT, ONNX, quantization, pruning, NVIDIA Jetson Orin (aarch64), ARM64 tuning
- **MLOps & Cloud:** Docker, Kubernetes, Cloud Run, Firebase Functions, CI/CD, MLflow, AWS Glue, SageMaker (familiar)
- **Data & Analytics:** ETL, 3NF & dimensional modeling, PostgreSQL (incl. vector DB usage), SQL, Tableau, Power BI
- **Languages & Frameworks:** Python, C++, Rust, SQL, Node.js, React, Qt

PROJECTS

ADVANCED SPORTS PERFORMANCE ANALYTICS PLATFORM

- Created advanced predictive models to extract actionable insights, including player positioning, performance metrics, and interactive dashboards for coaching staff and sports management.

FACE RECOGNITION SOFTWARE

- Developed a high-performance facial recognition system using MATLAB, leveraging transfer learning techniques with the AlexNet neural network, and achieved over 90% accuracy.

AI AGENT ECOSYSTEM FOR BUSINESS AUTOMATION

- Architected and developed a multi-agent system to automate various business departments including Sales, HR, Marketing, and Engineering.
- A central AI controller managed departmental agents, each responsible for 10-15 specialized tasks.

ADAPTIVE LANGUAGE LEARNING WEB APP

- Firebase + Gemini API; LLM-curated lessons, conversational practice, and progress tracking.

LOCAL RESEARCH AI AGENT

- Local-only agent using n8n + Ollama + SERP API; used PostgreSQL vector DB for knowledge retrieval and persistent memory.

PERSONAL ASSISTANT AI AGENT

- Developed a personal assistant agent using OpenAI's API and n8n to manage calendars, Gmail, Telegram, WhatsApp, and other personal tasks.

CLAUDE DESKTOP MCP CLIENT

- Created a local MCP client for Claude Desktop, enabling the AI to add, remove, and read files from the local machine.

ADVANCED DATA WAREHOUSE TRANSFORMATION

- Architected a 3NF data warehouse from 200M rows, implementing star schema designs for analytics and reporting, with ETL automation via AWS Glue and SQL

CLAIMBAI AGENT PLATFORM

- Architected multi-agent system integrating Google ADK, n8n, FastAPI, Firebase, and Cloud Run; implemented secure connectors and RAG-style retrieval.
- Built web app with AI agent connectors (email/calendar/tasks) and an automated questionnaire → recommendations pipeline.

EDUCATION

MIT MicroMasters, Statistics and Data Science (In Progress)	Virtual
Courses: Probability - The Science of Uncertainty and Data, Statistics and Data Science	
NUST	Islamabad, Pakistan
Bachelors, Mechatronics Engineering	

Certifications

- McKinsey Forward Program | McKinsey
- Tableau Data Analysis | Philip Burton
- Google Data Analytics | Coursera
- Advanced SQL for data analysis | Udacity
- UI/UX Design | Calarts
- CS50 | Harvard

Tools & Technologies

LLMs & Agents: LangChain, LlamaIndex, Ollama, n8n, Google ADK, local SLMs
ML / DL: PyTorch, TensorFlow, Hugging Face, YOLO, Whisper
Deployment / MLOps: FastAPI, Docker, Cloud Run, Firebase Functions, CI/CD, MLflow
Data & Infra: PostgreSQL (incl. vector DB usage), AWS Glue, ETL, Tableau, Power BI
Languages & Frameworks: Python, C++, Node.js, React, SQL, Qt