

AROON KUMAR

AI Engineer | +92 331 0390837 | Islamabad, Pakistan

aronk644@gmail.com | github.com/AroonKumarr | linkedin.com/in/aron-kumar0 | aron-kumar.vercel.app

SUMMARY

AI Engineer with 1 year of experience building and deploying production-grade Large Language Model (LLM) systems, Retrieval-Augmented Generation (RAG) pipelines, and Generative AI apps. Architected AigilityX, a full-stack AI platform integrating RAG, Speech-to-Text (STT), Text-to-Speech (TTS), and multi-agent orchestration. Built real-time voice AI using LiveKit and OpenAI Realtime API. Engineered a Redis-powered API Gateway reducing latency by 50% for 200+ concurrent users. Proficient in Python, LangChain, OpenAI API, FastAPI, Docker, and Redis.

SKILLS

AI & LLM	Python, LangChain, OpenAI API, GPT-4, Mistral, RAG, FAISS, Pinecone, Vector Databases, Prompt Engineering, Hugging Face, Multi-Agent Orchestration
Voice & Speech	LiveKit, STT, TTS, OpenAI Realtime API, Whisper, ElevenLabs
Backend & APIs	FastAPI, Flask, REST APIs, Redis, PostgreSQL, Supabase, WebSockets
DevOps & Tools	Docker, GitHub Actions, CI/CD, Railway, Git, AWS (EC2, S3, Lambda, Bedrock)

EXPERIENCE

AI/ML Solutions Executive Prudential Solutions <i>Part-time</i>	Aug 2025 – Feb 2026 Islamabad, Pakistan
---	--

- Promoted to part-time after 2 months — delivered production AI on AigilityX, a live Sovereign AI platform.
- Developed voice chatbots, multi-agent systems, and real-time Speech-to-Speech pipelines using LiveKit, STT/TTS.
- Engineered Redis-powered API Gateway reducing latency by 50% for 200+ users; deployed on Railway and Vercel.
- Architected ITP ConvertX — multi-city LLM chatbot for 5 cities, 3 languages, GPT-4 + Mistral fallback.
- Automated CI/CD via GitHub Actions, cutting manual deployment work by 70%.

Data Science Intern 10Pearls <i>Part-time</i>	Dec 2025 – Feb 2026 Islamabad, Pakistan
---	--

- Built AQI forecasting system using XGBoost and Scikit-learn achieving $R^2 > 0.85$ and MAE under 10% in production.
- Automated feature engineering, model training, and CI/CD daily retraining pipelines via GitHub Actions.
- Implemented Explainable AI (XAI) using SHAP and LIME for feature importance and model interpretability.

PROJECTS

Speech-to-Speech Voice AI. Real-time voice AI with STT → LLM → TTS pipeline using LiveKit Agents, OpenAI Realtime API, FastAPI token server, and Next.js frontend; deployed on AWS and Vercel. ([Live Demo](#))

AigilityX AI Platform. Full-stack Sovereign AI platform with RAG pipelines, STT/TTS services, multi-agent orchestration, Redis API Gateway, and Supabase backend; deployed on Railway. ([Live Demo](#))

AQI Predictor Islamabad. End-to-end ML forecasting using XGBoost and Scikit-learn with scheduled GitHub Actions pipelines, SHAP/LIME explainability, $R^2 > 0.85$, and live Streamlit dashboard. ([Live Demo](#))

EDUCATION

Bachelor of Computer Science , FAST-NUCES Islamabad, Pakistan	Feb 2022 – Feb 2026 CGPA: 3.56/4.0
---	---------------------------------------

Volunteer – Team Lead OGT, AIESEC Islamabad
Customer Experience (CXP) – managing EP documentation and exchange processes.