ZAINULLAH KHAN

AI Engineer | LLMs, RAG & Agentic Systems

Islamabad, Pakistan (Open to Remote) | zain.9496@gmail.com | +92 335 620 11 58
LinkedIn: linkedin.com/in/zainullah-k **| GitHub**: github.com/zainkhan-afk **| Website**: zainkhan-afk.github.io

PROFESSIONAL SUMMARY

AI Engineer with 7+ years building production LLM, NLP and computer vision systems. At SiParadigm, built RAG and agentic document-AI pipelines for healthcare that cut manual data entry by 80% and digitization cost by 50%. Previously Research Associate at NCAI (Urdu ASR) and Level 2 Fiverr seller with 400+ completed orders. MS Robotics and AI, NUST.

TECHNICAL SKILLS

LLM & Agents: LangChain, LangGraph, CrewAI, AutoGen, OpenAI Agents SDK, Gemini, Milvus, FAISS, pgvector
ML / CV / NLP: PyTorch, TensorFlow, Hugging Face, OpenCV, OCR, UMLS
AI-Assisted Development: Claude Code, OpenAI Codex, Cursor, GitHub Copilot, test-driven development
Backend & Data: Python, FastAPI, Django, Flask, Celery, SQL/PostgreSQL, Redis
Cloud & DevOps: Docker, GitHub Actions, AWS, Oracle Cloud, ONNX, TensorRT, vLLM, Ollama
Other: C++, React, JavaScript, TypeScript, Stripe, MFC, OpenGL

PROFESSIONAL EXPERIENCE

AI Developer II, *siParadigm Diagnostic Informatics* | Islamabad **June 2022 - Present**
US-based (New Jersey) specialty diagnostic laboratory providing molecular, genetic and pathology testing to physicians and healthcare providers.

- Built an asynchronous document digitization pipeline that uses batch LLM APIs, cutting digitization cost by 50% compared to the real-time APIs.
- Built a real-time agentic pipeline that converts scanned documents into structured digital records at $0.02-0.04 per page, using agents that extract document structure and hierarchy.
- Automated medical document parsing with OCR, LLMs and a healthcare-specific RAG pipeline, improving patient data-extraction precision and recall, and cutting manual data entry time by 80%.
- Built an LLM agent-based framework to detect and redact PHI in medical forms, achieving 90% precision and 95%+ recall in support of HIPAA compliance.
- Enabled structured clinical data modeling by building an entity relationship extraction pipeline using UMLS semantic types and relations.
- Achieved over 90% concept recognition accuracy in clinical documents by implementing medical text grounding with UMLS and MetaMap.
- Cut manual microscope calibration time by over 80% and improved image clarity by building an image-based autofocus system using edge detection and focal-stack analysis.
- Enhanced microscope imaging consistency across uneven sample surfaces by developing 3D plane interpolation logic using Delaunay triangulation for precise focus adjustment.
- Increased microscope sample throughput and minimized manual intervention by designing and integrating an autosampler module.

Freelancer (Level 2 Seller), *Fiverr* | part-time **Aug 2019 - March 2026**

Freelance AI engineering for international clients: computer vision, custom model training and deployment, and AI-powered automation.

- Completed 400+ orders for global clients across AI, computer vision, backend and automation, with a 4.9 rating.
- Built a real-time Twitch voice agent that transcribes streamer and chat audio, generates in-character replies with Gemini, and speaks through ElevenLabs text-to-speech.
- Built a social media automation SaaS web app (auth, Stripe subscriptions) with an ML model that predicts video virality, plus an automated post scheduler across Facebook, TikTok, Instagram, YouTube and multiple other platforms.
- Built Digital Media Bay, a SaaS web app (auth, Stripe subscriptions) that turns users' images and audio into animated vinyl-record videos.
- Created a Discord trading summary bot that automatically aggregates and summarizes daily, weekly, and monthly forex trading activity, improving trader productivity and performance reviews.

Research Associate, *National Center for Artificial Intelligence* | Islamabad **June 2021 - June 2022**

Pakistan's government-funded national AI research center, headquartered at NUST, building AI solutions for healthcare, media, judiciary and security.

- Trained, tested, and deployed an Urdu Automatic Speech Recognition (ASR) system, enabling accurate transcription for news channels.
- Improved content classification accuracy by building and deploying stream ad and program recognition models using audio and image processing and machine learning.
- Delivered insights on public opinion by performing topic-specific sentiment analysis on Twitter data using natural language processing techniques.

PROJECTS

Clipse - A shared clipboard across all your devices clipse-app.vercel.app

- A clipboard where you paste data from one device and copy it on another.

Physics Simulations and Creative Coding zainkhan-afk.github.io/Simulations

- Physics and creative coding simulations created using p5.js.

Research Paper RAG github.com/zainkhan-afk/Research-Paper-RAG

- A Retrieval Augmented Generation (RAG) project that answers questions from given research papers.

Differential Drive Robot MPC Controller github.com/zainkhan-afk/Differential-Drive-Robot-Navigation

- Differential drive robot navigation using Model Predictive Control.

OpenCV 3D Renderer github.com/zainkhan-afk/OpenCV-3D-Renderer

- A 3D renderer written from scratch using Numpy and OpenCV.

Robot Gait Visualizer github.com/zainkhan-afk/Quadrupedal-Robot-Gait-Visualization

- Quadrupedal robot gait visualization tool built on top of my OpenCV 3D Renderer.

EDUCATION

MS in *Robotics and Artificial Intelligence*, NUST – Islamabad 2020 - 2024

BS in *Electronic Engineering, BUITEMS* – Quetta 2015 - 2019