Abdullah Naeem

Lahore, Pakistan • +92 308 4488837 • abdullahnaeem2002a@gmail.com • LinkedIn • GitHub • Portfolio

EDUCATION

Forman Christian College University **Lahore, Pakistan**
BSc, Computer Science June 2026
GPA: 3.3
Relevant Coursework: Machine Learning, Deep Learning, DB Systems, Data Structures, Software Engineering, Web Development

EXPERIENCE

MIRAS AI **Lahore, Pakistan**
AI/ML Engineer Sep 2026 – Present
- Added Urdu to NVIDIA Nemotron 3.5 for low-resource-language voice agents, extending the tokenizer with a new Urdu token and fine-tuning on 121 hours of speech on a vGPU T40; the base model produced no Urdu output before fine-tuning.
- Reduced normalized WER to 22.1% and CER to 17.0% from a pre-fine-tuning baseline of effectively 100%, retaining 74.6% of English words in code-switched Urdu-English speech.
- Delivered streaming inference at 0.56s algorithmic latency (320ms chunk, 240ms lookahead) and 67ms p95 compute per chunk, reaching a 0.18 real-time factor that processes audio 5.5x faster than real time.

ARTHUS SYSTEMS **Remote**
AI/ML Engineer Jun 2026 – Aug 2026
- Built a real-time AI phone agent that answers intake calls for US law firms, using TypeScript and Fastify over a persistent WebSocket link to Retell AI and the Anthropic Claude API, replying in under 800ms end-to-end.
- Enforced each legal compliance step in code through a state machine rather than trusting the model to follow instructions, and screened callers for conflicts of interest mid-call in under a second using PostgreSQL fuzzy matching (pg_trgm).
- Added two layers of safeguards, prompt rules plus an output filter, to stop the agent giving legal advice it is not licensed to give, tested against 52 adversarial prompts in CI on every change.

TEACHING CIRCLE **Remote**
Software Engineer — AI Automation (Contract) Apr 2026 – Jul 2026
- Designed and deployed a production WhatsApp AI automation system for an education client, routing 100+ student enquiries/day on course details, payments and enrolment confirmations through OpenAI GPT-4 and GPT-5.1 mini orchestrated with LangChain.
- Built a webhook ingestion pipeline triggered by student form submissions, classifying enquiries at 95%+ accuracy across 100+ automated tests, with PostgreSQL logging and worker-level scaling that holds concurrent sessions without state collision.
- Automated the end-to-end enquiry workflow, cutting manual response workload by 80% and replacing a 2–3 person support team, containerised with Docker with no recorded outages in production.

PROJECTS

REAL-TIME DRIVER DROWSINESS DETECTION SYSTEM **GitHub**
Technologies: Python, PyTorch, TensorFlow, scikit-learn, OpenCV, MediaPipe, MobileNetV2
- Developed a 4-stage real-time CV pipeline using MediaPipe and MobileNetV2 with 60-frame temporal smoothing to suppress single-frame false positives, processing video at 40+ FPS on CPU with end-to-end latency under 50ms.
- Engineered a dual-model CNN for eye and yawn classification with two-phase transfer learning, reaching 99% accuracy on the MRL test set and 96% on YawDD against targets of 96% and 90%; verified attention with Grad-CAM.

PSYCHCONNECT: AI-POWERED TELEPSYCHOLOGY PLATFORM **GitHub**
Technologies: Next.js, FastAPI, Supabase, Gemini, Vertex AI, LangChain, FAISS, pgvector, Docker
- Built a RAG pipeline embedding clinical reference material into FAISS and pgvector, serving Gemini responses on Vertex AI to analyze projective drawings and PHQ-9/DASS-21 questionnaires and cutting assessment time by 20–30%.
- Built the platform and a Python matchmaking microservice in Next.js and FastAPI on Supabase, securing records with row-level security and role-based auth, delivered by a 3-person Agile team and validated by pilot testers.

SKILLS & CERTIFICATIONS

Programming: Python, TypeScript, JavaScript, SQL, C, R
AI & Machine Learning: LLMs, Generative AI, RAG, LangChain, Prompt Engineering, Embeddings, NLP, Computer Vision, Deep Learning, CNNs, Transfer Learning, PyTorch, TensorFlow, Scikit-learn, OpenCV, MediaPipe, NumPy, Pandas, Model Fine-Tuning, Speech Recognition (ASR), NVIDIA NeMo / Nemotron, Streaming Inference, Grad-CAM, OpenAI, Anthropic Claude API, LLM Evaluation, Red-teaming, promptfoo, Automation workflows
Platforms & Cloud: Google Cloud, Vertex AI, Gemini API, Docker, Vercel, Git, GitHub, Supabase, Vector Databases (FAISS, pgvector)
Backend & Infrastructure: REST APIs, FastAPI, Node.js, Fastify, WebSockets, PostgreSQL, Data Pipelines, OAuth 2.0, pytest, Agile, Scrum

Certifications: Machine Learning Specialization (DeepLearning.ai / Coursera, 2025) — Supervised Learning, Advanced Learning Algorithms, Unsupervised Learning, Recommenders and Reinforcement Learning