# Ahmad Makki

Lahore, Pakistan
+923042725542 makkia52@gmail.com linkedin.com/in/EngrAhmadMakki github.com/EngrAhmadMakki

## Experience

**Scraperrs Lab** **Feb. 2025 – Present**
*AI/ML Engineer* *Lahore, Punjab*
*Associate AI/ML Engineer*

- Worked on multiple computer vision projects, including model annotation using Roboflow, model training, testing, and evaluation.
- Developed FastAPI-based backends for inference and deployment of trained computer vision models.
- Containerized and deployed AI models using Docker, ensuring reliable and scalable integration into production environments.
- Handled client coordination and project reporting to align model performance with real-world requirements.
- Created an automated workflow on n8n for an Agentic AI system to streamline data and model operations.
- Built a conversational chatbot using OpenAI Whisper, large language models (LLMs), and in-context learning for intelligent voice and text interactions.
- Integrated NLP and computer vision solutions across two ongoing projects, combining multimodal AI capabilities for enhanced outcomes.

**Machine Learning 1 Limited** **Jan. 2025 – Feb. 2025**
*Trainee - Bootcamp Data Science & AI* *Lahore, Punjab*

- Developed and implemented machine learning models using Python libraries such as NumPy, Pandas, Scikit-learn, and SciPy.
- Performed data preprocessing for both supervised and unsupervised learning tasks, including data cleaning, feature extraction, and normalization.
- Applied Natural Language Processing (NLP) techniques, including tokenization, stemming, and vectorization, to train and optimize text-based models.
- Worked with senior data scientists to refine model performance, ensuring effective deployment and evaluation of models.

**Descon** **Oct. 2024 – Dec. 2024**
*Trainee - Data Science & AI* *Lahore, Punjab*

- Worked on skills in Machine Learning, Natural Language Processing (NLP), and AI under structured training.
- Completed the *Natural Language Processing with Classification and Vector Spaces* course by Coursera, gaining expertise in text classification and vector space modeling.
- Completed the *Generative AI with Large Language Models* course by Coursera, learning about transformer architectures, prompt engineering, and fine-tuning techniques.
- Completed the *Docker Mastery: with Kubernetes from a Docker Captain* course by Bret Fisher on Udemy, gaining hands-on experience with containerization, orchestration, and deployment strategies.
- Worked with NLP tools such as SpaCy and NLTK to preprocess text, extract features, and build classification models.
- Enhanced SQL proficiency for efficient data querying, manipulation, and database management.
- Engaged with CS50 coursework to strengthen computer science fundamentals and problem-solving abilities.

## Education

**Punjab University College of Information Technology - PUCIT** **Sep. 2023 – Present**
*MS Data Science (3.5 CGPA)* *Lahore, Punjab*

**University of Engineering and Technology - UET, Lahore**
*B.Sc Electrical Engineering*

## Projects

**Local AI Desktop Assistant with System Control (Electron MCP)** | *Python, LLaMA.cpp, MCP*

- Built a local AI-powered desktop assistant using LLaMA.cpp and Model Context Protocol (MCP), enabling fully offline chat, rewrite, summarize, and greeting functionalities.
- Implemented safe system-level controls including application launching, volume adjustment, brightness control, and power actions (lock/sleep), using strict whitelisting and intent-based routing.
- Designed a modular MCP server-client architecture allowing seamless integration with Electron/PyQt UI, global hotkeys, and extensible tool-based command execution.

**Real-time CV Inference API (YOLOv8 + FastAPI)** | *Python, YOLOv8, FastAPI, Docker*

- Deployed YOLOv8 object detection/segmentation models behind a FastAPI inference service for real-time predictions.
- Built production-ready REST endpoints for image/video inference with optimized pre/post-processing and structured JSON outputs.
- Containerized the service with Docker for reliable deployment and scalable integration into client applications.

**AI-Powered Portfolio Website (Interactive Demos + AI Assistant)** | *HTML, TailwindCSS, JavaScript, TensorFlow.js, Gem*

- Built an AI-powered portfolio with a futuristic UI (TailwindCSS) featuring dynamic sections, interactive timeline experience cards, and smooth navigation with responsive mobile support.
- Integrated in-browser Computer Vision demos using TensorFlow.js models (COCO-SSD object detection, MoveNet pose estimation, BodyPix person segmentation) running locally in the browser with real-time webcam overlays.
- Developed an embedded AI assistant using Gemini API with voice input (Web Speech API), text chat, and TTS audio responses, plus a Resume–JD match analyzer that outputs match score and missing keywords.

**Smart Conversational Agent for Mental Health Support** | *Python, ML, DL, NLP, LLMs*

- Developed a project on a Smart Conversational Agent to provide mental health support using various approaches.
- Implemented rule-based methods using CountVector, TF-IDF, and similarity matrices for initial response generation.
- Used retrieval-based techniques, including CNN, RNN, LSTM, GRU, and BiLSTM models, to retrieve appropriate responses based on user input.
- Incorporated generative-based models like GPT-2, GPT-3.5 Turbo, Cohere, MistralAi, and Ai21 Studio to create natural and empathetic conversations.

**Predictive Modeling for Student Final Scores** | *Python, ML & DL Models*

- Developed a predictive model to estimate students' final scores based on initial assessments.
- Trained the model using previous course data and tested it with data, focusing on the transition from initial to final.
- Utilized datasets from morning sessions for training and afternoon sessions for testing, predicting student scores from the 5th activity onward.

**Kubernetes ML Model Deployment Project** | *Minikube, Kubernetes, ML Models*

- Deployed an ML model using Minikube on a pod and demonstrated NodePort/LoadBalancer access.
- Implemented replica scaling, scaling the pod to 1-5 replicas and displaying the IP address of the serving pods.
- Tested CPU saturation, scaled to two pods, and implemented autoscaling in Kubernetes for performance optimization.

**Deployment of RAG-based PDF Query System** | *Docker, Streamlit, Deployment*

- Deployed a RAG-based (Retrieval-Augmented Generation) system that allows users to upload PDFs and query extracted information using a Streamlit-based interface.
- Containerized the application using Docker to ensure consistency across environments.
- Pushed the Docker image to Docker Hub for easy distribution and version control.
- Deployed the containerized application on a server, ensuring smooth access and performance.

## Certifications

**Coursera** **Fall 2023 – Present**
*IBM, AWS, DeepLearning.Ai* *Lahore, Punjab*

- Prompt Engineering by DeepLearning.Ai
- LangChain Chat with Your Data by DeepLearning.Ai
- Python for Data Science, AI & Development by IBM
- Deep Learning Specialization by DeepLearning.Ai
- Natural Language Processing with Classification and Vector Spaces by DeepLearning.Ai
- Generative AI with Large Language Models by DeepLearning.Ai

**Udemy**
*Docker, Kubernetes*

- Docker Mastery: with Kubernetes + Swarm from a Docker Captain by Bret Fisher

**HackerRank**
*SQL, Python*

- SQL Skills Certification
- Python Skills Certification

## Technical Skills

**Languages & Libraries**: Python, Pandas, NumPy, NLTK, SpaCy, Scikit-learn, LLMs (GPT-3.5 turbo, GPT-4, Cohere, MistralAi, and Ai21 Studio, LLAMA3)
**Developer Tools**: VS Code, Jupyter Notebook, Google CoLab, Git, Docker, Kubernetes
**Technologies/Frameworks**: Linux, GitHub, Matplotlib, Seaborn