

Muhammad Muaz Ashraf

AI/NLP Engineer

Contact

Lahore, Pakistan | +92 323-8479018 | muazashraf456@gmail.com | [LinkedIn](#): | [GitHub](#)

Education

BSc in Software Engineering | University of Lahore, Lahore, Pakistan | 2018 - 2023

Profile

AI Engineer with 2+ years of hands-on experience building NLP, RAG, and agentic workflows using OpenAI, LangChain, LlamaIndex, and Hugging Face. Expert in designing automation pipelines that reduce manual effort and enhance productivity. Proven success in remote roles, delivering production-grade solutions in fast-paced environments.

Tecnical Skills

Languages: Python

Frameworks & Tools: LangChain, LlamaIndex, LangGraph, Hugging Face, Flowise, LangFlow, Crewai, MCPs, MoviePy, OpenCV, Fireflies AI, ElevenLabs, Keyboard Maestro

AI/ML Techniques: RAG, Agentic RAG, Multimodal RAG, LLMs, Realtime transcription, TTS, STT, OCR

Databases: MySQL, SQLite, Vector DBs (Pinecone, Qdrant), Graph DBs

Other Tools: GitHub, Retool, Aider, Cursor, LiveKit, Twilio

Professional Experience

Junior NLP/Generative AI Engineer | Mojo Solo | Remote | March 2023 – Present

- Developed custom MCP servers deployed on DigitalOcean, improving internal workflow scalability and enhancing company productivity by 40%.
- Solved persistent issues in RAG vector database management using Flowise, significantly enhancing retrieval accuracy.
- Automated Midjourney image downloads via Discord and Keyboard Maestro, saving ~5 hours/week.
- Built a speech-to-text and text-to-speech system using OpenAI, ElevenLabs, and open-source models to streamline communication processes.
- Created a Vimeo video processing automation pipeline using MoviePy, saving 10 hours/week in manual editing.
- Engineered a parallelized web scraping tool, reducing data collection time by ~10 hours/week.
- Designed a RAG pipeline for video data using VideoDB and multimodal models, enabling precise content retrieval.
- Built agentic research tools capable of autonomous web search and summarization, reducing research response time.
- Implemented a Multimodal RAG using ColBERT, CLIP, and vector stores like Qdrant and Pinecone.
- Integrated Hugging Face BLIP for AI-generated storyboards.
- Developed a real-time transcript fetcher that extracts actionable insights.
- Implemented activity detection systems (e.g., pushups/squats) and face authorization using YOLO and OpenCV.

Certifications

Languages

- [Event-Driven Agentic Document Workflows, 2025](#)
- [Building AI Voice Agents for Production, 2025](#)
- [Career Essentials in Generative AI by Microsoft and LinkedIn](#)
- [Reinforcement Learning From Human Feedback, 2024](#)
- [Multimodal RAG: Chat with Videos!, 2024](#)
- [ChatGPT Prompt Engineering for Developers, 2023](#)

Urdu: Native
English: Professionally Proficient