MUNTAZIR ALI

AMBITIOUS, ENTHUSIASTIC, QUICK LEARNER

+92-3175160891 | 2019n03000@gmail.com | /muntazir-ali | Lahore, Pakistan



AI Engineer with 2+ years of hands-on experience designing and deploying AI-native systems using Generative AI, RAG pipelines, vector databases, and agentic frameworks. Strong background in Python backend engineering, scalable API development, and production AI deployments. Experienced in building chatbots, decision-making systems, and end-to-end ML workflows that power real-world applications.

SKILLS

- **Programming:** Python, PyTorch
- **LLMs/GenAI**: Livekit, Pipecat, STT, TTS, Gemini, Prompt Engineering
- **RAG/Agents:** LangChain, RAG Pipelines, Chunking, Reranking
- **Vector DBs**: Pinecone, FAISS
- **Backend:** FastAPI, REST APIs
- **Data/Pipelines:** ETL, Data Ingestion, PDF/Text Parsing
- **ML/NLP:** Transformers, Scikit Learn, Pandas, Matplotlib
- **Databases:** MySQL, MongoDB, Redis
- **DevOps/Cloud:** Docker, AWS
- **Tools**: Git, MLflow, Linux

COURSES

- IBM Machine Learning Specialization IBM, Oct 2024
- Applied Data Science with Python Specialization, University of Michigan, Jan 2025
- AI for Medicine Specialization, DeepLearning.AI , Jan 2025
- Python Mega Course Udemy, Aug 2024
- Aspire Leadership Program Harvard University, April 2024
- Data Structures University of California, San Diego, Aug 2024
- Advanced Algorithms and Complexity University of California, San Diego, Sep 2024
- Generative AI with AWS, June 2025
- Neo4j & LLM Fundamentals Neo4j, Dec 2024

HONOURS

- NIRD 2024 Winner
- Nominated for Prime Minister Laptop Scheme 2024.
- Secured a position at LUMS NOP Session 2019.

EDUCATION

Institute of Space Technology, Pakistan **2020 - 2024**

Bachelor of Engineering in Electrical Engineering Top 10%

Relevant Coursework: Programming Language, Probability in Engineering, Calculus, Linear Algebra, Embedded Systems, Computer Networks, Data Structures.

PROJECTS

AI Contact Center | Multi Voice AI Agents **February 2025**

- Built voice agent system handling 1000+ daily calls via SIP and WebRTC protocol.
- Reduced appointment no-show rates by 15% across 10 US practice through automated calls and also handling afterhours calls.
- Integrated medication refill and telemedicine workflows into unified platform.

Building Knowledge Graphs with LLM | Data Scientist **November 2024**

- Transformed 50K+ unstructured provider notes into Neo4j knowledge graphs using Transformer models.
- Achieved 92% entity extraction accuracy with RAG framework for patient diagnosis relationships.
- Enabled complex querying of clinical data relationships for treatment pattern analysis.

Deep Learning based-Brain Tumor Segmentation | Research Paper **July 2024**

- Designed novel encoder-decoder architecture improving Dice score by 5% over U-Net on BRATS 2020 dataset.
- Processed 3D MRI scans using PyTorch, optimizing GPU utilization by 40%.
- Implemented noise suppression techniques enhancing tumor boundary detection for 10,000+ medical images.

EXPERIENCE

CureMD **July 2024 - Present**

AI Engineer Lahore

- Deployed a production-grade multi-agent Voice AI platform for healthcare using SIP, Twilio, Asterisk, LiveKit, and FastAPI, supporting real-time inbound and outbound calls.
- Built low-latency speech pipelines integrating Deepgram for STT and ElevenLabs for expressive TTS, achieving natural, human-like real-time voice interactions with prompting techniques.
- Developed a full observability and analytics stack with Langfuse, enabling end-to-end LLM traceability, automated QA using LLM-as-judge, cost and latency monitoring, call summaries, success metrics, and workflow-level insights across 10+ practice use cases.
- Data intelligence systems using machine learning, Neo4j knowledge graphs, NLP, and GenAI to deliver actionable healthcare insights and improve clinical decision-making.

Qubit Dynamics **July 2024 - September 2024**

Bootcamp Remote

- Delved into tasks of STT model training, Gesture Recognition, Video Object Detection, Classification models, Emotion Recognition from Text.
- Explored Neural Network techniques like CNN, RNN and LSTM models.

PUBLICATIONS **Submitted**

Deep Learning-based Brain Tumor Segmentation