

MOHD FAIZY

AI/ML Engineer — LLMOps — GenAI — Agentic AI — Electronics Engineer

Email: mohdfaizy@gmail.com — Phone: +91-9891993929 — Location: New Delhi, India

[Portfolio](#) — [GitHub](#) — [LinkedIn](#) — [Twitter](#) — [Stack Exchange](#)

PROFESSIONAL SUMMARY

1+ years of experience building production-ready AI systems with expertise in **LLMOps, MLOps, and Deep Learning**. Proven track record deploying scalable AI applications using cloud-native technologies and modern DevOps practices. Strong Electronics Engineering background with specialized expertise in signal processing, computer vision, and generative AI.

CORE TECHNICAL EXPERTISE

AI/ML & Deep Learning

- **Frameworks:** PyTorch, TensorFlow, Keras, FastAI, DeepSpeed, ONNX, Scikit-learn
- **Specialized Libraries:** XGBoost, LightGBM, CatBoost, H2O.ai, OpenCV
- **Data Science Stack:** NumPy, Pandas, SciPy, Matplotlib, Seaborn, Plotly

Generative AI & LLMs

- **Models:** ChatGPT, GPT-4, Claude, Google Gemini, LLaMA, Mistral, BERT
- **Frameworks:** Hugging Face Transformers, spaCy, NLTK, OpenNLP, Rasa
- **Fine-tuning:** Model customization, transfer learning, prompt engineering

LLMOps & Production AI

- **Orchestration:** LangChain, LlamaIndex, LangGraph, AutoGen, CrewAI, Semantic Kernel
- **Vector Databases:** RAG, Pinecone, Chroma, Weaviate, FAISS
- **Deployment:** FastAPI, Docker, Kubernetes, AWS Fargate, GCP GKE
- **CI/CD:** Jenkins, GitHub Actions, GitLab CI, ArgoCD, CircleCI
- **Monitoring:** Prometheus, Grafana, ELK Stack, MLflow, Weights & Biases

Cloud & DevOps

- **Platforms:** AWS, Google Cloud, Microsoft Azure, Vercel, Railway, Linode
- **Containers:** Docker, Kubernetes (Minikube, GKE), Google Artifact Registry
- **Infrastructure:** Terraform, Infrastructure as Code, Auto-scaling
- **Security:** Trivy, SonarQube, Secret management, Security scanning

Databases & Web Development

- **Databases:** PostgreSQL, Neo4j, MySQL, SQLite, Vector Databases
- **Web Frameworks:** Django, Flask, FastAPI, Bootstrap
- **Frontend:** HTML5, CSS3, JavaScript, JSON, Streamlit

Development & Tools

- **Languages:** Python, C, LaTeX, Markdown, YAML, VHDL
- **IDEs:** PyCharm, VSCode, Cursor, Jupyter, Colab, Sublime Text
- **Version Control:** Git, GitHub, GitLab, Docker, Conda, Ubuntu
- **Design:** Canva, Adobe Photoshop

EDUCATION

M.Tech Electronics & Communication Engineering 2017-2019

Jamia Millia Islamia, New Delhi — CGPA: 8.6/10 (First Division with Distinction)
Thesis: IC Placement & Routing using Fully Convolutional Neural Networks

B.E Electronics & Communication Engineering 2012-2016

Jamia Millia Islamia, New Delhi — Score: 72% (First Division)
Project: Floating Point Multiplier & Simulation using VHDL

Diploma in Electronics 2008-2011

Jamia Millia Islamia, New Delhi — Score: 83% (First Division with Distinction)
Project: PIC Microcontroller Based Electronic Lock System

FEATURED PROJECTS

Production AI Systems

- **AI Anime Recommender:** Grafana Cloud monitoring, Minikube orchestration, ChromaDB vector storage, LangChain integration
- **Flipkart Product Recommender:** MLOps pipeline with Prometheus monitoring, Grafana dashboards, AstraDB integration
- **AI Travel Planner:** Kubernetes deployment with ELK stack monitoring, Filebeat logging, comprehensive observability
- **Study Buddy AI:** GitOps workflow with ArgoCD, Minikube orchestration, LangChain integration
- **Medical RAG Chatbot:** Production-ready system with Jenkins CI/CD, Trivy security scanning, AWS deployment, FAISS vector search
- **Radio Signal Classification:** Applied CNNs for multi-class space signal classification with production deployment
- **Facial Expression Recognition:** Multimodal emotion detection with Keras in production environment
- **Deepfake Detection:** Advanced GAN-based synthetic media detection system

Computer Vision & Object Detection

- **Real-Time Object Detection with YOLOv3:** High-performance real-time detection system
- **COVID-19 Detection:** Medical imaging analysis using PyTorch for chest X-ray diagnosis
- **Emotion AI & Facial Keypoints:** Advanced facial analysis with emotion recognition
- **Explainable AI Scene Classification:** GradCAM visualization for model interpretability
- **Traffic Sign Recognition:** Real-time CNN classification with TensorFlow deployment

NLP & Generative AI

- **GPT-2 Training from Scratch:** Complete implementation of transformer architecture
- **Recipe Generation with GPT-2:** Fine-tuned generative model for culinary applications
- **English-French Translator:** LSTM-based neural machine translation system
- **BERT Sentiment Analysis:** Advanced transformer-based sequence modeling
- **TensorFlow Model Serving:** REST API deployment with Flask integration

Machine Learning & Data Science

- **Linear Regression with NumPy:** Pure NumPy implementation of linear regression algorithms
- **Logistic Regression with NumPy:** Binary classification using NumPy from scratch
- **Multiple Linear Regression with scikit-learn:** Advanced regression modeling
- **Anomaly Detection in Time Series:** Keras-based temporal pattern recognition
- **Siamese Networks with Triplet Loss:** Advanced similarity learning in Keras
- **Imbalanced Data Classification:** Specialized techniques for skewed datasets
- **Support Vector Machine Classification:** Optimized SVM implementations
- **Support Vector Machines:** Advanced SVM implementations for classification tasks
- **Decision Trees:** Tree-based learning algorithms and ensemble methods
- **Autoencoder Image Denoising:** Deep learning approach to noise reduction
- **Dimensionality Reduction using Autoencoder:** Advanced feature extraction techniques
- **Simple Recurrent Neural Network with Keras:** RNN implementation for sequence modeling

Data Analysis & Visualization

- **Data Visualization with Plotly Express:** Interactive and dynamic data visualization
- **Statistical Data Visualization in Python:** Comprehensive statistical plotting and analysis
- **COVID-19 Data Visualization:** Interactive dashboards with Python and Plotly
- **World Map Geovisualization Dashboard:** Geographic data analysis and pandemic simulation
- **Principal Component Analysis with NumPy:** Dimensionality reduction implementation
- **Mining Data to Extract and Visualize Insights:** Advanced data mining techniques
- **Movie Recommendation System:** Collaborative filtering algorithms
- **Simulating Viral Pandemics in Python:** Epidemiological modeling and simulation

Specialized Applications

- **AI Sudoku Solver using AI and Python:** Constraint satisfaction and backtracking algorithms
- **Bioinformatics Tools:** Reverse and complement nucleic acid sequences (DNA, RNA) using Python
- **University Admission Prediction:** Multiple linear regression modeling for admission forecasting
- **Introduction to JavaScript The Basics:** Web development fundamentals and programming
- **Deploy Models with TensorFlow Serving and Flask:** Production model deployment pipeline
- **Serving TensorFlow Models with REST API:** Scalable model serving architecture

PROFESSIONAL CERTIFICATIONS

Latest Certifications (2024-2025)

LangChain LLM Applications - DataCamp — **Transformer Models & BERT** - Google Cloud

Data Science & ML Certifications

- **Associate Data Scientist:** DataCamp (Jun 2023)
- **Data Scientist Professional with Python:** DataCamp (Jun 2023)
- **PadhAI - Deep Learning Course:** One Fourth Labs (Aug 2022)

Specialized AI/ML Certifications

- **Deep Learning Specialization:** deeplearning.ai (Coursera)
- **TensorFlow Developer Certificate:** deeplearning.ai
- **AI for Medicine Specialization:** deeplearning.ai
- **Mathematics for Machine Learning:** Imperial College London (Coursera)
- **Advanced ML on Google Cloud:** Google Cloud Platform

Advanced Specializations

- **Natural Language Processing Specialization:** deeplearning.ai
- **Advanced Machine Learning Specialization:** Higher School of Economics
- **TensorFlow: Data and Deployment:** deeplearning.ai
- **Machine Learning with TensorFlow on GCP:** Google Cloud
- **Introduction to Deep Learning (with Honors):** Higher School of Economics
- **Practical Reinforcement Learning (with Honors):** Higher School of Economics

RESEARCH INTERESTS

- **LLMOps & Production AI Systems:** Scalable deployment and monitoring of large language models
- **Signal Processing:** Speech, audio, biomedical, and space signal analysis
- **Deep Learning for Healthcare:** Medical imaging, biomedical signal processing, and diagnostic AI systems
- **Multimodal AI:** Integration of vision, language, and audio processing for comprehensive AI solutions
- **Edge AI & Optimization:** Model compression, quantization, and deployment on resource-constrained devices

KEY ACHIEVEMENTS

- **Production AI Systems:** Successfully deployed 5+ enterprise-grade AI applications with 99.9% uptime
- **Performance Optimization:** Achieved 40% reduction in model inference time through advanced optimization techniques
- **Research Impact:** Published research on IC placement using FCNNs with significant performance improvements

- **Certification Excellence:** Earned 15+ professional certifications from top-tier institutions (Google, deeplearning.ai, DataCamp)
- **Open Source Contributions:** Active contributor to AI/ML community with multiple GitHub projects and technical articles

CONTACT INFORMATION

Available for full-time opportunities, consulting, and collaborative research projects