

ARSALAN SHAIKH

skarsalan8983@gmail.com | +91-8983665317 | [Portfolio](#) | [GitHub](#) | [LinkedIn](#)

PROFILE

Data Scientist with experience in OpenAI models tuning ,MLOps, medical imaging, financial risk analysis, and AI-driven platforms, focused on delivering reproducible models with real-world impact.

EXPERIENCE

Software Engineer Intern (Remote) | Dictation Daddy Sep 2025 -- Present

- Fine-tuned **OpenAI Whisper Large** for domain-specific transcription, improving accuracy in production use cases.
- Optimized inference pipeline, reducing **latency by 40%+** (1s → 600ms) for near real-time response.
- Deployed **scalable APIs** with monitoring, logging, and fault-tolerance for reliable integration.
- Integrated **Gemini 1.5 LLM** to deliver adaptive text formatting (formal/informal) based on context.

Machine Learning Intern (Remote) | Infosys Springboard Oct -- Dec 2024

- Developed and optimized a CNN model for Iris Eye Tumor Detection, achieving 92% accuracy.
- Processed and augmented 10K+ images, improving model reliability and reducing false positives by 18%.
- Built a scalable, real-time diagnosis pipeline with full project documentation and CI/CD integration.

PROJECTS

Aimers - AI Interview Platform (Demo) Jan -- Mar 2025

AI platform helping students assess knowledge and improve interview skills with personalized feedback

- Developed a BERT-based evaluation system achieving 90%+ accuracy against expert scores.
- Optimized FastAPI backend, reducing answer evaluation time from 1.2s to 300ms.
- Improved recommendation relevance by 40% using TF-IDF + cosine similarity and statistical analysis.
- Tracked experiments with MLflow and managed datasets with DVC for reproducibility.

SKILLS

- **Data Science / Machine Learning / Deep Learning:** Regression, Classification, Clustering, Ensemble Methods, PCA, CNN, RNN, LSTM, GRU, Model Tuning, Neural Networks, Artificial Intelligence, Feature Engineering, Feature Selection & Extraction
- **Natural Language Processing:** Preprocessing, Tokenization, Word2Vec, BERT, NLTK, spaCy, Text Classification, Similarity Analysis
- **Data Analysis & Visualization:** EDA, Visualization (Matplotlib, Seaborn, Power BI, Tableau), SQL (MySQL), NoSQL (MongoDB)
- **Mathematics for ML & DL:** Algebra, Probability, Statistics, Calculus, Matrices
- **Programming & Frameworks:** Python (NumPy, Pandas, SciPy, Scikit-Learn, TensorFlow, Keras, PyTorch, FastAPI), Data Structures, HTML, CSS, Flask, Figma
- **MLOps & Deployment:** CI/CD, Docker, MLflow, DVC, Render
- **Soft Skills:** Fast Learner, Problem Solving, Leadership, Optimistic, Communication Skills

EDUCATION

B.Tech in Artificial Intelligence & Data Science 2021 -- 2025
Marathwada Institute of Technology, Aurangabad