

# SHIKHER JAIN

Agra, Uttar Pradesh, India

✉ shikherjain786@gmail.com

📞 +91-7452017544

LinkedIn

Github

🌐 Portfolio

## Professional Summary

Data Scientist and AI/ML Engineer with hands-on experience in designing, developing, and deploying Machine Learning (ML) and Deep Learning (DL) models. Strong background in data preprocessing, exploratory data analysis (EDA), feature engineering, Natural Language Processing (NLP), and Computer Vision. Proficient in Python, PyTorch, and Scikit-Learn, delivering end-to-end AI solutions through production-ready, API-driven systems.

## Technical Skills

- **Programming Languages:** Python, Java, C
- **Machine Learning & Data Science:** NumPy, Pandas, Scikit-learn; Data Preprocessing, EDA, Feature Engineering
- **Deep Learning & AI:** PyTorch; Neural Networks, Model Optimization
- **Natural Language Processing (NLP):** Text Preprocessing, Tokenization, Embeddings, Transformer Models
- **Computer Vision:** OpenCV, MediaPipe, Face Recognition, Pose Estimation, Real-Time Vision Pipelines
- **Data Visualization & Analysis:** Matplotlib, Seaborn, Plotly, Tableau, Power BI
- **Web & API Development:** FastAPI, Flask, RESTful APIs, Streamlit
- **Databases & Tools:** MySQL, SQLite; Git, GitHub, Docker, Jupyter Notebook, Colab, Kaggle, Postman
- **Applied Mathematics & Logic:** Linear Algebra, Probability, Statistics, Rule-based Logic, Mathematical Modeling

## Work Experience

### Data Science Intern — Novas Arc Consulting Pvt. Ltd. (Remote)

Aug 2025 – Nov 2025

- Systematized an FAQ extraction pipeline for 100+ web pages, reducing manual effort by 60%.
- Built an NLP-based context analyzer to classify intent, tone, persona, domain, and age group, enhancing prompt relevance.
- Fine-tuned OpenAI GPT-3.5 Turbo on custom datasets, improving response relevance by 20–30% per internal evaluation.
- Deployed a training-data–driven AI chatbot with iterative testing and logic refinement, reducing inconsistent responses.
- Processed 10k+ text records for training and evaluation pipelines, ensuring high-quality model inputs.

## Education

### Faculty of Engineering and Technology, Agra College (AKTU)

B.Tech in Computer Science and Engineering

2022–2026

CGPA: 7.78/10

## Projects

### Sahayak AI— Multimodal AI Assistant

— Python, PyTorch, HuggingFace, FastAPI, Streamlit, GitHub, Live Qdrant, FAISS, SQLite, Docker

- Engineered a multimodal Retrieval-Augmented Generation (RAG) system supporting PDFs, images, audio, video, YouTube links, and text.
- Developed hybrid semantic search using Qdrant with automatic SQLite + FAISS fallback for reliable online/offline retrieval.
- Exposed RESTful APIs via FastAPI for embedding generation, semantic search, and summarization.
- Containerized PyTorch and HuggingFace models for seamless end-to-end inference using Streamlit.

### AttendEase — Smart Attendance System

— Python, OpenCV, NumPy, Pandas GitHub

- Developed a real-time face recognition attendance system using LBPH algorithm achieving 90% accuracy.
- Mechanized attendance logging into timestamped CSVs, reducing manual errors and improving operational efficiency.
- Replaced manual roll calls, optimizing attendance tracking and workflow.

### Census App — Interactive Data Visualization

— Python, Streamlit, Plotly, Pandas, NumPy GitHub, Live

- Built a Streamlit app to explore census data for 700+ Indian districts with dynamic filtering and interactive visualizations.
- Integrated Plotly-based maps and charts for actionable insights and faster data analysis.
- Improved data exploration efficiency by 40%, reducing manual work and enabling quick decision-making.

### Virtual Trainer — Real-Time AI Fitness Coach

— Python, OpenCV, MediaPipe GitHub

- Built AI fitness trainer using pose estimation to detect posture and count repetitions with 95% accuracy.
- Provided real-time feedback to reduce incorrect reps by 40%, leveraging MediaPipe holistic model at 30+ FPS.
- Enabled users to improve form and track performance metrics efficiently in real time.

## Certifications & Achievements

- TCS CodeVita Season 12, 2025 - Secured Global Rank 1040 out of 537,000+ participants worldwide ↗
- Event Coordinator - Cryptic Coder: Orchestrated campus workshops and coding contests
- Geodata Processing (Python & ML) - ISRO-IIRS, Feb 2025 ↗
- Machine Learning - Softpro India, AKTU, Jun 2024 ↗
- HP LIFE – Data Science & Analytics, Certificate of Completion, Dec 2024 ↗
- Python Programming - DataFlair, May 2024 ↗
- Python - HackerRank, Apr 2024 ↗