Harsh Sharma

harsh.sharma4131@gmail.com | 9767755630 | GitHub | LinkedIn

Professional Summary

Pre-final year AI & ML Engineering student with a strong passion for building intelligent systems that solve real-world problems. Hands-on experience developing LLM-based RAG applications, deep learning classifiers, and end-to-end ML pipelines. Aspiring Machine Learning Engineer driven by curiosity and a builder's mindset – from medical chatbots to brain tumour detection, I turn ideas into working models. Actively seeking opportunities to contribute, grow, and push the boundaries of applied AI.

Technical Skills

Languages & Libraries: Python, C, NumPy, Pandas, Scikit-learn, PyTorch, Matplotlib, Seaborn, Streamlit
Deep Learning: LSTM, CNN, Embedding Layers, Tokenization, RAG, LLMs, Transfer Learning
Tools & Databases: Git, GitHub, Jupyter Notebook, VS Code, MySQL, MongoDB

Experience

Zidio Development 08/2025 – 10/2025
Data Scientist Intern

- Developed an Image Captioning system leveraging Deep Learning (LSTM, Embedding layers, CNN features) to generate descriptive captions for images.
- Performed Exploratory Data Analysis (EDA) on image-text datasets to identify patterns, clean data, and optimize preprocessing pipelines.
- Implemented tokenization, sequence padding, and word embeddings for preparing captions for training neural networks.

Projects

Medical Chatbot – LLM-Based RAG System | *Python, LangChain, Pinecone, Streamlit*

- Built an end-to-end Retrieval-Augmented Generation (RAG) chatbot to provide accurate, context-aware information about medicines and drug interactions.
- Integrated Pinecone vector database for efficient semantic search and retrieval over a medical knowledge base of 10,000+ documents.
- Deployed interactive Streamlit interface allowing users to query medicine details, dosage, and side effects with source-cited LLM responses.

Brain Tumour Classification | *Python, PyTorch, CNN, Transfer Learning, Streamlit*

- Designed and trained a CNN-based deep learning model to classify brain MRI scans into four categories: glioma, meningioma, pituitary tumour, and no tumour.
- Applied transfer learning using a pre-trained VGG16 architecture, achieving 95%+ classification accuracy on the test dataset.
- Built a Streamlit web application enabling users to upload MRI images and receive instant classification results with confidence scores.

Sentiment Analysis of Movie Reviews | *Python, LSTM, Keras, NLP*

- Developed a deep learning model using LSTM architecture to predict movie review sentiment with 90% accuracy on a 50,000-review dataset.
- Applied text preprocessing techniques including tokenization, stopword removal, and sequence padding for optimal model performance.

Education

GH Raisoni College of Engineering and Management Pune, Maharashtra
Bachelor of Technology, Computer Science and Engineering (AI & ML) *2023 – 2027*
Sanjay Ghodawat Institute Maharashtra
Higher Secondary *2023*

Certifications

Cloud Computing – Infosys (July 2024)
Generative AI – IBM (November 2024)
Introduction to Data Science – Cisco (September 2025)