VISHAL KUMAR

Brookefield, Bengaluru – 560037.

+91 8789728908 | vishalgsu@gmail.com| linkedin.com/in/vishal-gsu | github.com/Vishal-gsu

PROFESSIONAL SUMMARY

Software Engineer specializing in Artificial Intelligence with demonstrated expertise in building and deploying machine learning models. Proficient in Deep Learning (CNNs), Computer Vision (OpenCV), and Generative AI, with hands-on experience in architecting Retrieval-Augmented Generation (RAG) systems and real-time inference pipelines. Seeking to leverage skills in TensorFlow, Python, and model optimization to develop scalable, high-impact AI solutions.

EDUCATION

B.E - Artificial Intelligence & Machine Learning (2022-2026)
CMR Institute of Technology, Bengaluru, CGPA: 7.6

TECHNICAL SKILLS

- **Programming & Core CS:** Python, Java, REST APIs, Git, GitHub
- **AI/ML Frameworks:** TensorFlow-Keras, PyTorch, Scikit-learn, LangChain, OpenCV
- **AI Specializations:** Machine Learning, Deep Learning (CNNs), Computer Vision, NLP, Generative AI, RAG
- **Data Visualization:** Tableau, Power BI
- **Data & Vector Databases:** FAISS, ChromaDB, Supabase, SQL
- **Development & Automation:** FastAPI, Streamlit, n8n, Docker
- **Development Tools:** Generative AI tools (ChatGPT, Gemini CLI), AI-assisted code editors (Cursor), App development(Expo)

PROJECTS

Scout Vision: AI-Powered Blind Assistive System

- Engineered a full-stack, AI-powered assistive technology system to provide real-time environmental awareness for visually impaired users.
- Architected a client-server system featuring a React Native mobile app for real-time data capture and a powerful, laptop-hosted Python server for complex AI inference.
- Implemented low-latency, real-time object detection by deploying a pre-trained **YOLOv8** model on a FastAPI backend, processing video streams from the client.
- Tech stack: Expo, React-native, Python, FastAPI, PyTorch, YOLOv8, Model Fine-Tuning, Quantization, LangChain, Gemini Pro, Git.

Offline RAG Chatbot

- Engineered a fully offline **Retrieval-Augmented Generation (RAG) chatbot** for semantic querying of private PDF documents.
- Implemented a data processing pipeline for text extraction and vectorization using a **local embedding model (all-MiniLM-L6-v2)**, storing indices in a **FAISS vector database** for efficient retrieval.
- Developed a **Streamlit-based UI** with persistent session history and integrated conversational memory to support multi-turn, context-aware dialogue.
- **Tech Stack:** LangChain, Ollama (Mistral), FAISS, Sentence-Transformers, Streamlit, PyPDF, PyTorch.

Deep Learning System for Real-Time Sign Language Interpretation

- Designed a **low-latency sign language recognition** engine, leveraging a custom-trained neural network to classify and translate complex hand gestures from a real-time video stream with high accuracy.
- Engineered a sophisticated feature extraction pipeline using **MediaPipe** to process video frames and generate high-dimensional, 3D spatial coordinate data representing hand landmarks for model ingestion.
- Deployed a **Multi-Layer Perceptron (MLP) model**, trained on the extracted landmark data, to perform robust, multi-class classification of signs, achieving minimal inference time for a seamless user experience.
- **Tech Stack:** Python, TensorFlow, Keras, OpenCV, MediaPipe, Scikit-learn, NumPy.

Agentic AI for Deep Research

- Developed a **multi-Agentic AI architecture** for deep research. Allowed customizable research depth, structure, and clarifying questions for personalization. Agent stack designed to handle layered research tasks of generating SERP queries, scrape web results, generate learnings and append it to a stack of knowledge. Displayed results with citations. Implemented Web scraping, authorization, backend automation, and database access.
- **Tech Stack:** Python, LangChain, Supabase, Vector DB, Google-OAuth.

ACHIEVEMENTS

- Analyzed a complex dataset and presented actionable business insights using Tableau in a Deloitte virtual experience on Forage.