# Soumyashis Sarkar

Hyderabad, Telangana
8369994800 | soumyashissarkar362@gmail.com
LinkedIn: https://www.linkedin.com/in/soumyashis-sarkar/ | GitHub: https://github.com/Chitragupta16

## Education

### Vellore Institute of Technology

*Bachelor of Technology in Computer Science and Engineering* *Oct 2022 – Aug 2026*

- CGPA: 8.46/10.0

## Work Experience

### Machine Learning Intern **July 2024 – Apr 2025**

*Omdena* *Remote*

- Led team of 5 members in Bhopal BRTS chatbot development, collected 300+ route data points using web scraping and manual methods, deployed Streamlit web application
- Implemented geospatial routing algorithms using Haversine formula and OSRM API integration, built RAG pipeline with LangChain and Gemini API for dynamic query responses
- Collaborated with 15+ international engineers via Slack and Google Colab across Frankfurt green space mapping using Sentinel-2 satellite imagery and VM-UNet architecture

## Projects

**LLM Hallucination Prevention System** | *Python, Streamlit, FastAPI, NLP* **Nov 2025**

- Engineered ML-powered fact verification pipeline extracting 7 entity types from LLM responses using spaCy NER, cross-referencing against Wikipedia API with 90%+ accuracy for high-confidence claims
- Designed contradiction detection algorithm tracking factual consistency across conversation history, implementing similarity scoring and numeric variance analysis to flag conflicting statements
- Deployed production-ready web application on Streamlit Cloud with RESTful API backend, featuring side-by-side response comparison, color-coded verification indicators, and automated citation generation

**Player Re-Identification in Soccer Matches** | *Python, YOLOv11, OpenCV, Deep SORT, PyTorch* **Jun 2024**

- Developed player detection and tracking pipeline using YOLOv11 for bounding box detection and Deep SORT for consistent ID assignment
- Implemented re-identification logic to maintain player identities across occlusions and overlapping frames in real-time video feeds
- Simulated match scenarios ensuring accurate tracking for analytics, with robust handling of ball-player interactions

**Urban Green Space Mapping using VM-UNet** | *Python, PyTorch, CUDA, OpenCV, Docker* **Dec 2024 – Mar 2025**

- Adapted VM-UNet architecture from medical to geospatial applications for multi-channel Sentinel-2 satellite imagery segmentation
- Implemented data augmentation techniques improving model performance: F1-score increased to 82%, accuracy to 77.4%, specificity to 77.9%
- Built training pipeline using PyTorch with CUDA optimization, deployed on Google Colab with containerized environment setup

## Technical Skills

**Programming Languages:** Python, C++,Java, SQL, JavaScript
**Framework:** PyTorch, TensorFlow, Node.js
**Development Tools:** Git, VS Code, Kaggle, AWS, Docker

## Extracurriculars

- Finalist at Smart India Hackathon 2024 with gamified Indian Consitution educational platform
- PR & Outreach Member at AI Club, VIT Bhopal
- Team Lead, Freelancing Club, VIT Bhopal