HIMANSHU GANGWAR

+918954628679 | gangwarsanu8@gmail.com
Himanshu Gangwar | zenitsu0509 | Portfolio
Baheri, UttarPradesh - 243201, India

Experience

- **Whissle AI** *Mar 2024 – Jul 2024*
  AI/ASR Intern California (Remote)
  - Achieved a 12% Word Error Rate (WER) reduction in production ASR models by fine-tuning transformer architectures on diverse acoustic datasets, enhancing transcription accuracy in noisy environments.
  - Developed a scalable data preprocessing pipeline using Python and PyTorch, processing over 50,000 hours of raw audio with automated quality validation, reducing model training cycles by 30%.
  - Benchmarked SOTA ASR architectures (e.g., Conformer, Wav2Vec 2.0), leading to the adoption of an optimized model that decreased inference latency by 25% and improved out-of-vocabulary (OOV) accuracy.

Projects

- **Arya Bhatt Hostel AI Agent**
  Tech: LangChain (ReAct), RAG, Groq API, Pinecone, Streamlit, Sentence-Transformers
  - Developed an intelligent AI agent using the LangChain ReAct framework with function-calling to automate hostel management tasks like complaint handling, info retrieval, and administrative queries 24/7.
  - Engineered a RAG pipeline with 100+ documents embedded in a Pinecone vector database, enabling low-latency LLM responses and dynamic tool selection via the Groq API.
  - Planned a production-ready Streamlit app with persistent chat memory, serving the hostel community by automatically resolving queries on mess schedules, complaints, and more.
- **MEDGraphy: Graph RAG Drug Information App**
  Tech: Graph RAG, Neo4j, LangChain, Groq, Llama3, Streamlit, Sentence-Transformers
  - Architected a Graph RAG system using a Neo4j knowledge graph with 10,000+ nodes (drugs, diseases, side effects) to provide context-rich answers to complex medical queries.
  - Implemented a hybrid search combining Cypher queries and semantic vector search, feeding enriched context into a Llama3-8B model via the Groq API for real-time, accurate response generation.
  - Revitalized an interactive Streamlit app with graph visualization, allowing users to receive text answers while exploring connections between medical entities.
- **Automatic Number Plate Recognition (ANPR) System**
  Tech: YOLOv8, PyTorch, CRNN, OpenCV, Gradio, Python
  - Engineered an ANPR pipeline combining a YOLOv8 detector (98% mAP) and a PyTorch-based CRNN OCR model (99.2% character accuracy) for robust performance in varied conditions.
  - Fine-tuned deep learning models on a custom dataset, achieving high-precision plate detection and character recognition through a two-stage computer vision architecture.
  - Deployed a real-time Gradio web application with <200ms end-to-end processing for immediate and interactive license plate recognition.

Education

- **Institute of Engineering and Technology** *Sept 2023 – Sept 2027*
  B.Tech in Computer Science and Engineering (AI) Lucknow, India
- CGPA: 7.9/10.00

Skills

- **Languages** Python, C++, C, Java
- **AI/ML** PyTorch, TensorFlow, Scikit-learn, LangChain, Transformers, RAG, ASR, LLMs, OpenCV, Sentence-Transformers
- **Databases** Vector DB (Pinecone, FAISS), Neo4j, MySQL, Pandas
- **Web/Deployment** Streamlit, Gradio, Docker, Render, Vercel, Git, GitHub

Certifications & Achievements

- **Key Achievements:** Top 1% in Uttarakhand Board 2023 (Class 12); Winner of college-level hackathon.
- **Certifications:** AWS Machine Learning Foundations, Google Cloud (Prompt Design in Vertex AI, MLOps with Vertex AI), DeepLearning.AI (AI Agents in LangGraph).