

HIMANSHU GANGWAR

+918954628679 | gangwarsanu8@gmail.com

 Himanshu Gangwar |  zenitsu0509 |  Portfolio

Baheri, UttarPradesh - 243201, India

EXPERIENCE

• Whissle AI

Mar 2025 - Jul 2025

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

• Employee NLQ: Production-Grade Natural Language Query System



Tech: React, FastAPI (Python), TypeScript, PostgreSQL, Vector DB (pgvector/FAISS), LLMs (Groq/OpenAI), Redis, Docker

- Engineered a full-stack NLQ platform with auto-schema discovery, relationship inference, and hybrid query execution across structured (SQL) and unstructured (document) data sources without hard-coded schemas.
- Implemented dual SQL generation: rule-based heuristics for common HR queries and LLM-powered (Groq Llama/OpenAI) translation for complex JOINS, aggregations, and safe DML (UPDATE/INSERT/DELETE) with transaction handling and affected row tracking.
- Built scalable document ingestion pipeline: multi-format support (PDF/Word/CSV/Excel/JSON), intelligent chunking by document type, sentence-transformer embeddings, and configurable vector stores (in-memory FAISS or persistent pgvector).
- Deployed production features: Redis-backed async job queue, tabular data import with chunked loading, TTL-based response caching, query history tracking, and comprehensive Material-UI dashboard with real-time job status polling.

• 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.

• 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

SGPA: 8.3/10.00 · 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, Hugging Face |
| • Databases | PostgreSQL, MySQL, Vector DB (Pinecone, FAISS), Neo4j, Pandas |
| • Web/Deployment | FastAPI, Streamlit, Gradio, Docker, Render, Vercel, Git, GitHub |
| • Cloud/DevOps | Google Cloud, Microsoft Azure |

CERTIFICATIONS & ACHIEVEMENTS

- Winner:** College hackathon at Anukuran Fest (Dec 2024); Led a team to build a PDF-to-Mindmap generator. 
- Key Achievements:** Top 1% in Uttarakhand Board 2023 (Class 12)
- Certifications:** AWS Machine Learning Foundations, Google Cloud (Prompt Design in Vertex AI, MLOps with Vertex AI), DeepLearning.AI (AI Agents in LangGraph).