

# TEJAS GARG

+91-8815548460 | [tejas@tejasgarg.in](mailto:tejas@tejasgarg.in) | [tejasgarg.in](http://tejasgarg.in) | [github.com/garg-tejas](https://github.com/garg-tejas)

## EDUCATION

### Indian Institute of Information Technology

B.Tech. Computer Science & Engineering (AI & ML) | CGPA: 8.32/10

Nagpur, Maharashtra

August 2023 – June 2027

## TECHNICAL SKILLS

**Languages:** Python, TypeScript, JavaScript, SQL, C/C++

**Speech & LLM Systems:** Whisper (faster-whisper), Wav2Vec, Deepgram, HuggingFace Transformers, LLM Fine-Tuning (GRPO, LoRA, SFT), Prompt Engineering, Structured Outputs

**RAG & Retrieval:** LangChain, LangGraph, Hybrid Search (BM25 + Dense + RRF), Cross-Encoder Reranking, FAISS, Vector DB Integration

**Backend & Systems:** FastAPI, Flask, Async IO, PostgreSQL, Redis, SQLAlchemy, REST APIs

**MLOps & Deployment:** Docker, vLLM, AWS Bedrock, Vercel, Supabase, Mixed Precision (AMP)

## PROJECTS

### SalesLens: Real-Time Sales Call Intelligence | *Whisper, Pyannote, FAISS, FastAPI, Redis*

 [GitHub](#)

- Built an end-to-end voice AI pipeline: faster-whisper STT transcription, pyannote speaker diarization, and LLM-based role identification to convert raw call audio into structured, speaker-attributed transcripts.
- Designed an LLM analysis layer extracting objections, action items, call scores, and sentiment per speaker turn; integrated FAISS-powered RAG for contextual Q&A over conversation history.
- Engineered async FastAPI backend with PostgreSQL and Redis for pipeline orchestration and state management.
- Routed inference through HuggingFace Router (Qwen 3.5-9B) to balance latency and cost across multilingual call data.

### CoreCS Interview Lab | *Hybrid RAG, FastAPI, PostgreSQL, LangChain*

 [GitHub](#)

- Built a production RAG pipeline indexing 9 CS textbooks with hybrid retrieval (BM25 + dense embeddings + RRF) and cross-encoder reranking, delivering citation-grounded LLM responses via GLM 4.7 over Z.AI API.
- Designed an offline LLM evaluation pipeline (Generation → Scoring → Schema Validation → Seeding) reducing malformed outputs from 12% to <1% through strict JSON guards and retry logic.
- Implemented prerequisite-aware learning paths using topological sort with SWOT/mastery signals, combined with SM-2 spaced repetition stored in PostgreSQL for adaptive review scheduling.
- Architected a FastAPI service layer with SQLAlchemy, Alembic migrations, and a React/TS/Vite frontend for a full-stack production AI system.

### ArXiv Literature Scout | *LangGraph, Multi-Agent, FastAPI, LangSmith*

 [GitHub](#)

- Architected a LangGraph multi-agent system with shared state management across Discovery, Analysis, and Survey subgraphs for automated research workflows.
- Integrated Semantic Scholar, arXiv, and Firecrawl APIs with per-agent LLM routing (gpt-oss-120b for reasoning, Qwen 3.5 for generation) to balance cost and output quality.
- Implemented 4 human-in-the-loop checkpoints and structured output retry logic, reducing malformed responses to near zero; traced all agent hops via LangSmith.

### GRPO Fine-Tuning on Mistral-7B (GSM8K) | *HuggingFace, LoRA, vLLM, PyTorch*

 [GitHub](#)

- Fine-tuned Mistral-7B Instruct with 4k SFT examples followed by Group Relative Policy Optimization (GRPO) to induce structured chain-of-thought reasoning traces.
- Improved GSM8K test accuracy from 41.17% (zero-shot) to 52.46% (+8% absolute over SFT baseline) on 2.9k RL training samples.
- Optimized training under 24GB VRAM using LoRA, gradient accumulation, and mixed precision; accelerated inference via vLLM batched decoding.

## RELEVANT COURSEWORK

**Core CS:** Data Structures & Algorithms, Design & Analysis of Algorithms, OOP, DBMS, OS, Computer Networks

**AI/ML:** Machine Learning, Artificial Intelligence, Natural Language Processing, Computer Vision

**Mathematics:** Linear Algebra, Probability & Statistics, Discrete Mathematics

## CERTIFICATIONS

Machine Learning Specialization – DeepLearning.AI (Andrew Ng)

Deep Learning Specialization – DeepLearning.AI (Andrew Ng)

AWS Future AI Engineer – Udacity