Tejas Garg

+91-8815548460 | contact.tejasgarg@gmail.com | linkedin.com/in/tejas-garg | github.com/garg-tejas

Education

Indian Institute of Information Technology — Nagpur, Maharashtra
B.Tech. Computer Science & Engineering (AI & ML) | ***CGPA: 8.32/10*** — *Aug. 2023 – June 2027*

Technical Skills

Languages: Python, C++, TypeScript, JavaScript, SQL
ML/Deep Learning: PyTorch, scikit-learn, Hugging Face Transformers, YOLO, OpenCV
Frameworks & Libraries: FastAPI, Next.js, React, Flask, Tailwind CSS
Databases & Tools: PostgreSQL, SQLite, Redis, Docker, Git, Linux
Cloud & DevOps: AWS Bedrock, Vercel, REST APIs

Projects

Explainable AI for Diffusion-Based Medical Classifiers | ***PyTorch, DiffMIC-v2, XAI*** GitHub

- Reproduced DiffMIC-v2 from scratch — dual-granularity conditional diffusion classifier achieving 84.1% accuracy on APTOS-2019 diabetic retinopathy dataset, matching the original paper
- Reduced memory footprint 10x via mixed-precision training and gradient checkpointing, enabling training on single GPU under 15GB VRAM
- Built custom XAI framework with 6 explainability techniques for diffusion models: temporal trajectory analysis, attention maps, faithfulness validation, and counterfactual steering that flips predictions with ¡3% pixel perturbation
- Designed full-trajectory attribution system backpropagating through 1,000 timesteps, revealing 78% of severe DR predictions stem from local lesion features

SentinelVision: Real-Time PPE Compliance Monitor | ***FastAPI, YOLOv11, SAM3, Next.js*** GitHub

- Built event-driven video processing system converting noisy frame-level detections into stable violation events with start/end times and evidence snapshots
- Implemented async pipeline with decoupled display (30 FPS) and ML inference (3 FPS) using producer-consumer pattern to prevent frame queue buildup
- Applied temporal filtering with EMA confidence fusion and hysteresis thresholds, reducing false positive alerts by 60% under occlusion
- Designed REST API for event querying and statistics with SQLite async persistence; deployed full-stack with Next.js dashboard

NL-to-SQL: Interpretable Query Generation Pipeline | ***Flask, Qwen, HuggingFace*** GitHub

- Built multi-stage LLM pipeline: schema parsing → chain-of-thought reasoning → SQL generation → verification with auto-correction
- Achieved 78% execution accuracy on Spider benchmark with transparent reasoning steps for human auditability
- Implemented prompt injection detection and input sanitization layer to prevent SQL injection attacks
- Added automatic retry with error recovery for LLM timeouts; exposed type-safe REST API with Pydantic validation

GRPO: Teaching Mistral-7B to Reason | ***PyTorch, Unsloth, vLLM*** GitHub

- Trained Mistral-7B with Group Relative Policy Optimization on GSM8K, achieving 52.5% test accuracy (+11% over zero-shot) through induced XML-formatted reasoning traces
- Discovered critical evaluation bug where structured outputs were penalized vs. free-form, demonstrating importance of unified correctness functions in RLHF
- Implemented full SFT→GRPO pipeline on consumer GPUs (2×T4) using LoRA quantization and vLLM for 10x faster evaluation

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