Pranav Anand

Chennai, Tamil Nadu | 8778497231 | pranavofficial.acc@gmail.com | pranavoffi.github.io/Pranav_portfolio

Professional Summary

Technically skilled B.Tech student with hands-on experience in **Large Language Models**, **data analysis**, and **interactive AI systems**. Proficient in PyTorch, LangChain, and LangGraph for image generation, sentiment analysis, and **RAG-based retrieval**. Skilled in **LLM fine-tuning** and building scalable, production-grade AI pipelines for real-world impact.

Work Experience

Junior AI Engineer Intern Dec 2025 – March 2026
Mako IT Lab

- Architected a dual-agent **Sales Intelligence** system using **LangGraph** — a deterministic Sales Agent pipeline (document extraction → query generation → retrieval → pitch synthesis) paired with a context-aware RAG Chatbot with conversation-aware query rewriting
- Built a **hybrid retrieval pipeline** on **Qdrant Cloud** combining dense embeddings, sparse vectors, and a **cross-encoder reranker** with parent-child chunking to retrieve contextually relevant document sections
- Integrated **LangExtract** for schema-based structured extraction from unstructured company profiles; generated multi-angle business queries per prospect to maximize recall across industry and strategic dimensions
- Deployed a **Streamlit** frontend with real-time LLM token streaming, session persistence, and tool integrations including **Firecrawl** web scraper and **BrightData** Twitter scraper; instrumented the full pipeline with **OpenTelemetry** tracing and custom latency metrics

Technical Skills

Programming Languages: Python, SQL, HTML, CSS

Frameworks & Libraries: PyTorch, TensorFlow, LangChain, LangGraph, FastAPI, Streamlit, OpenCV, Unsloth, LangExtract, HuggingFace

Databases & Vector Stores: MySQL, SQLite, PostgreSQL, ChromaDB, Qdrant, Supabase, Neo4j, Redis

AI/ML Specializations: LLMs, Generative AI, Hybrid Retrieval (Dense + Sparse), GraphRAG, Agentic RAG, Multi-agent Systems, LLM Fine-tuning (LoRA/QLoRA), SLM Inference Optimization, Cross-encoder Reranking, Parent-child Chunking, RAG Evaluation (RAGAS), Knowledge Graph Construction, Conversation Memory Management, Prompt Engineering, GANs

System Design: LLM Pipeline Architecture, Production RAG Deployment, Session Management & Persistence, LLM Token Streaming, Request Queuing (Redis/Celery), Observability & Latency Metrics, MCP (Model Context Protocol), Agent-to-Agent (A2A) Communication, Docker

Tools & APIs: Firecrawl, BrightData, OpenAI API, Groq API, Anthropic Claude API, OpenTelemetry, LLM Guard/Guardrails, mem0, Web Scraping & Enrichment

Key Projects

AI-Powered Shoe Design Generator | *LLaMA, Stable Diffusion, LangChain* July 2025 – Aug 2025

- Built a generative AI assistant that **combines LLMs and image models** to transform abstract user conversations into context-aware shoe designs; **fine-tuned LLaMA 3.2 3B** to generate highly accurate prompts for Stable Diffusion v2
- Scalable solution for the footwear industry spanning automated concept design to personalized retail applications

Satellite Image Translation using Pix2Pix GAN | *GANs, U-Net, Streamlit* June 2025

- Built Streamlit-based deployment with custom-trained **U-Net generator** and **PatchGAN discriminator** for real-time satellite imagery translation into map-style renderings
- Achieved **model stability** and **visual coherence** in complex satellite imagery translation; designed for large-scale GIS and urban planning applications

StyleAI – Personalized Fashion Recommendation System | *YOLOv8, VLM, RGCN, FastAPI* 2025

- Built a 5-stage AI pipeline using **YOLOv8 segmentation** to split user photos into face and body crops, analyzed by fine-tuned **VLMs** (LLaVA 7B + Gemma 3) to build a **PhysicalProfile**, fused into a 512D user state vector via 5 parallel MLP encoders
- Scored recommendations using a 2-layer **RGCN** Fashion Knowledge Graph with live product crawling; powered by **Llama 3.1 8B** multi-turn conversational engine over WebSocket streaming, deployed via **FastAPI**

Education

Bachelor of Technology in Computer Science Engineering Expected 2026
SRM Institute of Science and Technology ***CGPA: 8.15*** *Chennai, Tamil Nadu*

Specialization in Artificial Intelligence and Machine Learning

Senior Secondary (XII) – CBSE Percentage: **74.4%** July 2022

Secondary (X) – CBSE Percentage: **72.6%** July 2020

Achievements

HackForge'25 National Hackathon – Top 6 Finalist June 2025
National Level Competition *India*

- Achieved top 6 finalist position at national-level hackathon focused on developing AI-powered scalable solutions