

Vedant Shelkar | IIT Madras, India

Chennai-600036, Tamil Nadu, India | +91 8625074280 | vedantshelkar@gmail.com | [GitHub](#) | [LinkedIn](#)

Education

Indian Institute of Technology (IIT) Madras | B.Tech. | CGPA: 8.15/10

2023-2027

Experience

AI Developer Intern | GoLLM Limited, Scotland, UK

Dec'25 – Jan'26

- Built an end-to-end automated historical survey analytics pipeline with **advanced ML techniques** (including **clustering** and **statistical tests**) & **Time series** analysis across temporal dimension enabling richer, cross-sectional insights.
- Reduced the **analysis time by 10%** through parallelized workflows with **LangGraph** integrated with **Celery** asynchronous task distribution ensuring horizontal scalability and resilient task processing.
- Engineered a token-efficient API orchestration layer, **reducing API calls** by approximately **25%** through batching and payload optimization, **cutting API costs** by **40%** supporting high-throughput processing of large-scale AI inference workloads.
- Architected scalable **cloud data infrastructure** with **AWS S3** (raw data lake) and **MongoDB** (indexed query layer) with schema version control and evolution, yielding high-performance real-time survey analytics.

Research Intern | Guide: Prof. Murali Jagannathan, IIT Madras

Apr'25 - Aug'25

- Engineered **NLP pipeline** using **Tesseract OCR** and **spaCy** to automate legal document parsing and accelerate data processing.
- Finetuned spacy en_core_web models for **NER** & content traction modules for clause classification & parsing for structured document analysis.
- Evaluated model performance using **BLEU** and **ROUGE** metrics, achieving a **40%** improvement over base en_core_web and an **F1 score of 86.1%** against human-annotated ground truth letters
- Reduced manual claim & **BoQ (Bill of Quantities)** preparation time **by up 90%** through **advance NLP integration** and targeted **model fine-tuning**

IICPC Pvt. Ltd. | Operations and Management Intern, Chennai

Dec'24 - Apr'25

- Spearheaded the deliverable execution and partnerships with **5 leading HFT firms**, overseeing deliverable execution across **10+ tasks**, resulting in **100% on-time delivery** and strengthened long-term partnerships.
- Executed and optimized end-to-end logistics for **IICPC Codefest across 8 IITs in India**, managing **45,000+** participants and reducing execution time by **30%** through process streamlining and team coordination.
- Directed financial operations for **INR 30L+** in partner deliverables, improving ROI for sponsors through customized branding strategies.

Projects

Conversational Analytics

Sept'25

- Developed a conversational AI data analytics platform using **LangGraph (DAG)** and **FastAPI** (Async REST API) with modular router architecture and **Pydantic** schema validation, orchestrating multi-file analytics with LLM-driven planning, LLM code generation and intelligent retry mechanisms.
- Built a **Celery** (async task queue) + **Redis** (brokering and caching) **async distributed task pipeline** integrated with **PostgreSQL** (SQLAlchemy async ORM) to process concurrent large-scale datasets via decoupled parallel analysis workers & query level result **caching** without blocking API response cycles.
- Containerised the platform using Docker** and implemented a multi-layer security framework for executing LLM-generated Python code for dangerous pattern blacklisting (eval/exec/import injection), infinite loop detection, **SHA-256 file deduplication** and redaction via **regex** credential sanitizer.

Finance Bot

Apr'25

- Architected a **Streamlit**-based chatbot integrating **Llama 3.1** via **Ollama** to analyse & compare financial documents (PDFs), enabling automated extraction of key insights using large language models (LLMs)
- Developed **Agentic RAG** system with layout-aware system using **LangChain** & **LangGraph**, built vector database with **Pinecone DB** with scalable **Indexing strategies** leveraging layout segmentation via **LayoutLMv2** (transformer based) & **Detectron2** (**Faster R-CNN** and **RetinaNet**) for visual backbone processing ensuring structured output from unstructured financial doc for financial insights from income statements, balance sheets & MD&A reports.
- Orchestrated a data pipeline with **PDFPlumber**, **FastEmbeddings** & strategic text chunking with **RecursiveCharacterTextSplitter** to process large documents and detect red flags, benchmark performance metrics & interpret trends

Medical Query

Feb'25

- Fine-tuned transformer-based **GPT-2** model with **PEFT** (Parameter-Efficient Fine-Tuning) **QLoRA** achieving **88 % domain accuracy**
- Enhanced text generation quality using advanced sampling strategies (**temperature=0.7**, **top_k=50**, **top_p=0.9**) and a repetition penalty of **1.2**.
- Achieved 45% memory savings** in training by leveraging **QLoRA** and **low-bit quantization** techniques with **PyTorch** and **Hugging Face**.

Prediction Modelling

Nov'24

- Developed **Random Forest** & **XGBoost Ensemble** on 27 years of UEFA data to predict winners on compressed feature space through PCA
- Achieved **79.9%** balanced accuracy using **PCA (98% variance)** & hyperparameter tuning with **GridSearchCV** on engineered input features.
- Generated explainable insights through **SHAP analysis**, **PCA loadings**, and feature importance visualization, leveraging **Seaborn** and **Matplotlib**

Technical Skills

Machine Learning & Deep Learning: Finetuning (SFT, RLHF, IFT), PyTorch, TensorFlow, LangChain, LangGraph, HuggingFace Transformers, Scikit Learn, Cuda
Programming & Tools: Python, C++, Git, Github, PostgreSQL, Docker, Celery, FastAPI, Redis, Pandas, SQL, NumPy, AWS, GCP, MongoDB, Qdrant, HTML5

Achievements

VentureSprint | E-Summit'25, IIT Madras

Feb'25

- Ranked 2nd nationally among 300+participants for creatively solving one of ten cross-sector challenge statements
- Proposed a **"Tap & Pay"** credit card activation model using **SIM-style activation technology**, driving a 40% boost in adoption across **Tier II/III regions**.
- Brought down **Customer Acquisition Cost (CAC)** by **55%** & onboarding time by **85%**; was offered a **pre-incubation opportunity** by **PedalStart**

Positions of Responsibility

Sponsorship and PR Coordinator | Shastra 2025

May '24 – Apr'25

Shastra is India's largest student-run technical fest with **50K+ footfall** having received **ISO 2015 certification**

- Raised **INR 2.2 Lakhs** in sponsorships from a key corporate partner, contributing to a **10% Year-over-Year (YoY)** increase in sponsor count and coordinated cross-team executions of **INR 1 Cr+** in sponsor deliverables across **15+** companies including **Boeing, Caterpillar, Macmillan and Digikey**