

## EDUCATION

B.Tech in Computer Science Engineering (8.33/10.0)  
Minor in International Relations, Specialisation in Data Science

Aug'23 - July'27

## EXPERIENCE

LLM Intern, Critical AI Pvt Ltd, Bengaluru

July'25 - Current


- **GenAI Audio & Image based narrative creation using Meta's Llama-3.2-8B**
  - Integrated a multimodal GenAI system with RAG, Whisper STT, gTTs, image search, and OSM mappings.
  - Business Impact: Improved functionality by over 50% of base provided console.
  - Software tools: llama-cpp-python, faiss, PyMongo

AI Intern, WESEE, Ministry of Defence, New Delhi


May'25 - July'25

- **Data Analysis and Report Generation Agent with MCP architecture**
  - Built a robust chatbot pipeline for processing real-time SQLite data, and generating reports for real-time usage.
  - Business Impact: Reduced the workload of critical subject matter experts by more than 90%.
  - Software tools: Langchain, SQLAlchemy, LangGraph, mcp-use, Ollama
- **YOLOv8 Object Detection and Tracking with automated selenium web-scraped datasets**
  - Delivered object detection and tracking system, tuned on custom webscraped dataset, receiving over 75% accuracy.
  - Business Impact: Improved reliability of real-time usage by more than 50%.
  - Software tools: Ultralytics, Selenium

## PROJECTS

**XLM-RoBERTa based Multi-Lingual sentiment analysis with auditory input support** 


- Developed pipeline that applies multi-modal analysis of audio input, with 5 supported languages scoring 70% accuracy.
- Utilises whisper for audio-to-text transcription, XLM-RoBERTa for textual sentiment analysis, AudioCNNs for auditory sentiment analysis of MFCCs and an MLP model for data fusion.
- Technology Stack: PyTorch, scikit-learn, FastAPI, transformers

**GPT-2 Style LLM creation and Meta LLaMA-3.2 8B LoRA fine-tuning for Data Science workloads** 

- Built GPT-2 Architecture, replicating key components like MHA, and implementing KV-Cache for 5x performance.
- Fine-tuned Meta LLaMA-3 8B employing PEFT training using LORA, and quantized it to Q5K-M as a GGUF.
- Technology Stack: PyTorch, TikToken, Unsloth, Transformers, PEFT, HuggingFace

**XGBoost based Churn Prediction with SHAP Analysis and powerBI dashboard** 

- Designed feature engineering, training and deployment pipeline for telecom churn: to understand why, and take action to reduce it, with powerBI dashboard and streamlit frontend for intractability, and SHAP analysis with 70%+ scores.
- Includes data preprocessing, XGBoost training, Optuna hyperparameter tuning, and multiscored evaluation.
- Technology Stack: scikit-learn, shap, xgboost, optuna

**Hybrid Anime Recommendation System with CI/CD pipeline and Kubernetes deployment** 

- Deployed pipeline to process data, train NCF model, and deploy on GKE, with over 73% reliability for recommendations.
- Used Jenkins for orchestrating CI/CD pipeline, tracked experiments with comet-ml, flask for API creation, and Google Cloud Registry for handling docker images, and DVC for data versioning.
- Technology Stack: comet-ml, flask, tensorflow, dvc

## TECHNICAL SKILLS AND CERTIFICATIONS

Languages	Python, C, C++, SQL
Libraries and Frameworks	Pandas, Numpy, PyTorch, scikit-learn, mlFlow, llama-cpp-python, Jenkins, kubernetes
Certifications	MLOps with GCP - CI/CD & Kubernetes – KrishAI Technologies, 2025

## ACHIEVEMENTS

- Dean's list: Monsoon'23
- Semifinalist speaker at SMVIT Parliamentary Debate'24 (Debating Society, )