Kaisar Imtiyaz

ikaisar10@gmail.com kaisarimtiyaz kaisarimtiyaz mrbane10

Indian Institute of Technology, Kharagpur *Dec '21 – May '25*
B.Tech in Mechanical Engineering *CGPA : 7.70/10*

Skills and Technologies

Programming & Infrastructure
Python, SQL, Git, VS Code, Linux, Bash, Docker, AWS

Data Science & ML
Numpy, Pandas, Scikit-learn, SciPy, PyTorch, TensorFlow

Applied AI & NLP
OpenCV, NLTK, SpaCy, Hugging Face, FAISS, Haystack

Frameworks & Deployment
FastAPI, Streamlit, LangChain, LlamaIndex, LangGraph

Experience

IntegriSphere — AI Engineer *Internship — Remote — Sep '25 – Oct '25*
- Built an ***Agentic RAG*** based Chat Assistant enabling survey inspectors to query unstructured facility data from the O&G industry
- Designed a ***Deep Research Agent*** for automated analysis and report generation from ***Cathodic Protection*** and ***NDT*** records
- Implemented a hierarchical ***LangGraph*** workflow with parallelly spawned researchers managed by a supervisor through orchestration

Gupshup — Machine Learning Engineer *Internship — Saket, Delhi — Jun '24 – Sep '24*
- Contributed to adapting open source LLMs like ***LLaMA 3*** for scalable multilingual customer support in low-resource Indic languages
- Applied ***Instruction Fine-tuning*** with multiple chat templates to precisely tailor LLMs' responses for customer support interactions
- Employed techniques like ***LoRA*** and ***PEFT*** to obtain ***95***% parameter reduction, lowering training compute and memory overhead
- Achieved a ***20% perplexity*** reduction and ***10% BERTScore*** gain across languages, boosting multilingual fluency and consistency

Imago AI — Machine Learning Developer *Internship — Remote — Dec '24*
- Worked on ***Hyperspectral Imaging*** models for mycotoxin detection and predicting concentration levels in crops like corn and wheat
- Developed predictive models, starting with a classical ***XGBoost*** model and then a ***CNN-LSTM*** architecture for improved accuracy
- Contributed to improving ***codebase reproducibility*** and ***debugged*** existing models to reduce error rates and improve runtime

Projects

Fraudulent Résumé and Endorsement Detection System *Innov8 2.0, IIT Delhi, Eightfold AI — Sep '24*
- Engineered a system to identify résumé fraud and uncover suspicious endorsements in recommendation letters by detecting falsifications
- Applied ***Leiden clustering*** via ***NetworkX*** for anomaly detection, identifying fraudulent patterns and circular endorsement schemes
- Designed novel metrics including Loyalty Index, Redundancy Score, and Exaggeration Score to quantify fraud likelihood in résumés
- Created an HR-facing dashboard for fraud insights securing ***5th place*** out of 250 competing teams in the Innov8 2.0 competition

English-to-Urdu Transformer Architecture from Scratch *Self — Jul '25*
- Implemented and pretrained a ***64M*** parameter Transformer based model from scratch using PyTorch, for English-to-Urdu translation
- Processed OPUS-100 with WordLevel tokenization, added special tokens, and applied causal masking for ***autoregressive decoding***
- Trained on Kaggle's P100 GPU for **20 epochs** using Adam optimizer; training loss dropped from ***8.0*** to ***1.8***, tracked via ***WandB***
- Achieved BLEU ***0.44***, CER ***0.36***, WER ***0.53*** on test set; built custom inference pipeline and deployed model on ***Hugging Face Hub***

Semantic Search and Q&A System for Lecture Videos *Bachelor's Thesis I — Prof. Saud Afzal*
- Automated transcript ingestion from YouTube/NPTEL and built topic-content embeddings for a searchable course knowledge base
- Formulated a weighted topic-content similarity ranking for retrieval and leveraged **Groq** for low-latency, context-grounded responses
- Deployed the app on Streamlit with model switching and LaTeX support, delivering $<$ ***1s*** lookup and $<$ ***2s*** end-to-end response time

RAG-Based Chatbot for Engineering Textbooks *Bachelor's Thesis II — Prof. Saud Afzal*
- Developed a RAG chatbot for engineering PDFs, handling files up to ***200MB*** on CPU/GPU with concurrent multi-session management
- Applied conversation-aware ***query disambiguation*** and top-k vector retrieval to optimize context extraction for technical queries
- Deployed on Streamlit with Groq's inference, LaTeX rendering, metadata source linking, and debug visualization for relevance scoring

Multilingual Intent Detection Using Fine-Tuned BERT *Indo ML 2024, IIT Bombay*
- Fine-tuned BERT on the Amazon MASSIVE dataset (***1M+*** utterances, ***52*** languages, ***60*** intents) for multilingual intent classification
- Engineered custom tokenization pipelines and PyTorch datasets, leveraging HuggingFace's ***Trainer API*** for efficient model training
- Achieved ***89% accuracy*** and an ***F1 score*** of ***87.5***%, demonstrating robust performance across both precision and recall metrics

Certifications & Coursework

Certifications : Python 3 Specialization (UMich), Fundamentals of Statistics (MITx), Machine Learning Specialization (Stanford) Deep Learning Specialization (Stanford), Probability & Statistics for ML (DeepLearning.AI), LLMs Course & AI Agents (HuggingFace)
Coursework : Programming & DS, Linear Algebra, Calculus (Adv. & Transform), PDEs, Operations Research, Robotics, Soft Computing

Academic Achievements

- Ranked in the ***top 1%*** of 1 million candidates in ***JEE Mains***. Achieved ***top 5%*** ranking among 200k candidates in ***JEE Advanced***
- Secured a highly competitive ***department change*** at IIT Kharagpur by performing in the ***top 5%*** among 1,800 first-year students